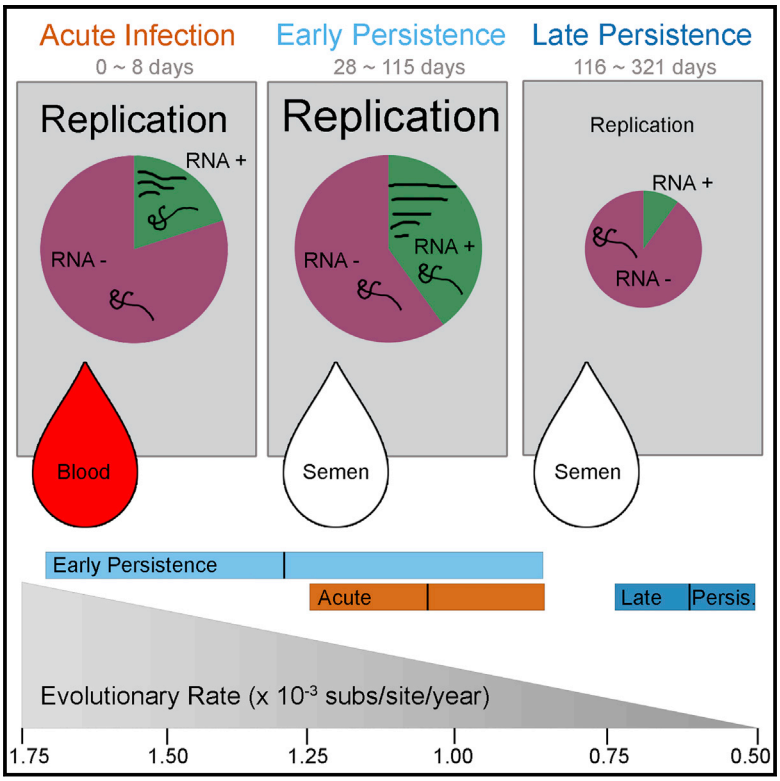
Report

Cell Reports

Active Ebola Virus Replication and Heterogeneous Evolutionary Rates in EVD Survivors

Graphical Abstract



Authors

Shannon L.M. Whitmer, Jason T. Ladner, Michael R. Wiley, ..., Gustavo Palacios, Ute Ströher, Ebola Virus Persistence Study Group

Correspondence

evk3@cdc.gov (S.L.M.W.), ute.stroeher@gmail.com (U.S.)

In Brief

Whitmer et al. find that Ebola virus continues replication/transcription within the eye and male genital tract of Ebola virus disease survivors. They describe viral replication, evolutionary rates, and selective pressures experienced during acute and persistent infection.

Highlights

- During persistence, EBOV exhibits heterogeneous evolutionary rates
- Active EBOV transcription and replication occurs during persistence
- RNA hyper-editing observed during viral persistence
- No evidence for significant selective pressure during persistence

Data and Software Availability

KY401638–KY401675
KY805810-2

Whitmer et al., 2018, Cell Reports *22*, 1159–1168
January 30, 2018
https://doi.org/10.1016/j.celrep.2018.01.008

CellPress

Cell Reports
Report

Active Ebola Virus Replication and Heterogeneous Evolutionary Rates in EVD Survivors

Shannon L.M. Whitmer,[1,11,14,*] Jason T. Ladner,[2,11,12] Michael R. Wiley,[2,11] Ketan Patel,[1] Gytis Dudas,[3] Andrew Rambaut,[4,5] Foday Sahr,[6] Karla Prieto,[2] Samuel S. Shepard,[7] Ellie Carmody,[8] Barbara Knust,[1] Dhamari Naidoo,[9] Gibrilla Deen,[10] Pierre Formenty,[9] Stuart T. Nichol,[1,11] Gustavo Palacios,[2,11] Ute Ströher,[1,11,13,*] and Ebola Virus Persistence Study Group

[1]Viral Special Pathogens Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA
[2]Center for Genome Sciences, US Army Medical Research Institute of Infectious Diseases, Frederick, MD, USA
[3]Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[4]Institute of Evolutionary Biology, University of Edinburgh, King's Buildings, Edinburgh, UK
[5]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
[6]Sierra Leone Armed Forces, Freetown, Sierra Leone
[7]Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA
[8]Division of Infectious Diseases, NYU School of Medicine, Bellevue Hospital Center, New York, NY, USA
[9]Health Emergency Programme, World Health Organization, Geneva, Switzerland
[10]Sierra Leone Ministry of Health and Sanitation, Freetown, Sierra Leone
[11]These authors contributed equally
[12]Present address: Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, USA
[13]Present address: Independent Researcher, Mannheim, Germany
[14]Lead Contact
*Correspondence: evk3@cdc.gov (S.L.M.W.), ute.stroeher@gmail.com (U.S.)
https://doi.org/10.1016/j.celrep.2018.01.008

SUMMARY

Following cessation of continuous Ebola virus (EBOV) transmission within Western Africa, sporadic EBOV disease (EVD) cases continued to re-emerge beyond the viral incubation period. Epidemiological and genomic evidence strongly suggests that this represented transmission from EVD survivors. To investigate whether persistent infections are characterized by ongoing viral replication, we sequenced EBOV from the semen of nine EVD survivors and a subset of corresponding acute specimens. EBOV evolutionary rates during persistence were either similar to or reduced relative to acute infection rates. Active EBOV replication/transcription continued during convalescence, but decreased over time, consistent with viral persistence rather than viral latency. Patterns of genetic divergence suggest a moderate relaxation of selective constraints within the sGP carboxy-terminal tail during persistent infections, but do not support widespread diversifying selection. Altogether, our data illustrate that EBOV persistence in semen, urine, and aqueous humor is not a quiescent or latent infection.

INTRODUCTION

From December 2013 to June 2016, Sierra Leone, Guinea and Liberia experienced an Ebola virus (EBOV) outbreak causing 28,646 confirmed, probable, and suspected Ebola virus disease (EVD) cases—including 11,323 deaths and over 10,000 EVD survivors (WHO, 2016a). Despite the World Health Organization (WHO) declaring these countries disease-free 42 days (twice the 21-day viral incubation period) after the last active case, sporadic EVD cases continued to appear outside of this window and several reports strongly suggest that these unexpected re-emergences occurred due to viral transmission from persistently infected EVD survivors (Arias et al., 2016; Blackley et al., 2016; Christie et al., 2015; Diallo et al., 2016; Mate et al., 2015; Sissoko et al., 2017b). Other possible explanations, later discarded, included that sporadic cases could represent a missed transmission chain, reintroduction from an animal reservoir, or from another geographical location. Genetic data and phylogenetic analysis have been critical toward a resolution among these possibilities.

Filovirus persistence was initially observed with a single Marburg virus sexual transmission case in 1967 (Martini and Schmidt, 1968). Very scarce data from previous outbreaks suggested a prolonged presence of EBOV nucleic acids in semen and other bodily fluids collected from convalescent patients (Bausch et al., 2007; Rodriguez et al., 1999; Rowe et al., 1999). Recent EVD persistence studies in Sierra Leone, Liberia, Guinea, and the United States extended these observations and definitively demonstrated that EBOV RNA can be detected within the semen of EVD survivors months to ~2 years after recovery, and live virus can be isolated from a subset of these specimens (Barnes et al., 2017; Deen et al., 2017; Sissoko et al., 2017a; Soka et al., 2016; Uyeki et al., 2016). Initially, the WHO and Médecins Sans Frontières (MSF) advised male survivors to abstain from sexual intercourse or use barrier protection for 3 months after recovery (Sterk, 2008; WHO, 2014), however, based on results from the current outbreak (Christie et al.,

2015; Deen et al., 2017; Mate et al., 2015), the WHO revised their recommendations to include periodic EBOV RT-PCR semen testing and for survivors that cannot access EBOV RT-PCR semen testing, they should continue to practice safe sex for at least 12 months after the onset of symptoms (WHO, 2016b). Viral recrudescence outside of the male genital tract (MGT) can also develop after filovirus infection, as initially observed in 1977 for a single case of Marburg virus uveitis (Kuming and Kokoris, 1977). During the West African outbreak, recrudescent cases were again observed within the eye, and also the CNS several months after initial infection (Jacobs et al., 2016; Varkey et al., 2015). Altogether, these data suggest that after recovery from EVD, EBOV can still persist within immune-privileged sites in EVD survivors.

While much work has been done to explore the molecular evolution of EBOV during acute infection (Dudas et al., 2017; Gire et al., 2014; Ladner et al., 2015; Park et al., 2015; Simon-Loriere et al., 2015; Tong et al., 2015), little is known about the dynamics of persistent EBOV infections within immune-privileged niches. Genomes from EVD flare-ups linked to transmission from persistent infections exhibited reduced genetic divergence (Blackley et al., 2016; Diallo et al., 2016). These low levels of divergence could help to define and predict whether new outbreaks are the result of transmission from individuals with acute or persistent infections—such data could influence and guide future epidemiological investigations. Furthermore, the extraordinary discovery that EBOV can persistently infect immune-privileged sites for several months opens significant questions regarding viral replication mechanisms and the selective pressures experienced during acute and persistent infection.

To address these questions, we directly sequenced EBOV RNA from clinical specimens collected during acute EVD and during EVD convalescence ("persistence") (Figure S1A). Using these EBOV sequences, we directly estimated viral evolutionary rates during persistent infection. We observed significant reductions in the rate of viral evolution within a subset of persistent infections, while others exhibited acute-like rates, and we present potential mechanisms to explain these results. We also examined patterns of selection during persistent infection and demonstrate that active viral replication/transcription continues during viral persistence.

RESULTS

EBOV in Semen Specimens from Sierra Leonean EVD Survivors Exhibits Reduced Evolutionary Rates

Using a random subset of acutely acquired viral sequences (AAVS) from specimens collected from May 2014–September 2015 and sequenced directly from blood, plasma, or oral swab specimens from EVD patients with acute symptoms in Sierra Leone, we inferred a mean evolutionary rate of 0.96×10^{-3} substitutions/site/year ([$0.86–1.06 \times 10^{-3}$] 95% credible interval) under the uncorrelated lognormal (UCLN) model of rate variation among branches. These acute rate estimates are similar to previously reported rate estimates (Gire et al., 2014; Park et al., 2015; Simon-Loriere et al., 2015; Tong et al., 2015). Using Bayes factor values calculated from path and stepping-stone sampling, the UCLN relaxed clock models were the best fit to the data, however, evolutionary rate estimates were also similar under the fixed local clock model (Figure S1B; Table S1).

Most semen-acquired viral sequences (SAVS) exhibited lower genetic divergence, given their sampling time, than the mean AAVS divergence, although in all cases, this divergence was inside the prediction interval calculated for AAVS (Figure 1A). The average collection period for SAVS was 170 days post disease onset with a range of 82–322 days. During these collection periods, SAVS exhibited a significantly reduced evolutionary rate compared to AAVS (Figure 1B). Reversion of potential U-to-C hyper-edited sites, which may be the result of host ADAR enzymes, similar to Dudas et al. (2017), slightly decreased the acute evolutionary rate (0.89×10^{-3} subs/site/year, [$0.80–0.99 \times 10^{-3}$] 95% credible interval), as expected (Figure 1A). After removal of hyper-edited sites, SAVS exhibited a marginally significant reduced evolutionary rate compared to AAVS (Figure 1B).

EBOV Evolutionary Rates from Paired Acute and Convalescent Clinical Specimens

Serial specimens acquired from US EVD survivors permitted a comparison of viral sequences acquired during acute and persistent infection within a single individual. For all US survivors, AAVS were nearly identical and exhibited genetic divergence consistent with other AAVS collected during the outbreak (Figures 2A and S1C). For survivor C, concurrent viral compartmentalization was observed in the eye and MGT, and we did not observe evidence of viral exchange between these sites (Figure S1C). Using the UCLN relaxed clock model, mean posterior rate estimates from US AAVS (estimated over an average of 5 days) were slightly decreased, but not significantly different to rate estimates from other AAVS collected during the outbreak (estimated over 542 days) (Figure 2B; Table S1). In contrast to SAVS collected from EVD survivors in Sierra Leone, SAVS collected from US EVD survivors exhibited a mean evolutionary rate estimate that was ~1.45-fold greater than acute rate estimates (Figure 2B; Table S1). We attribute this rate increase to U-to-C hyper-editing that occurred during viral persistence in survivors A and C (Figures 2A and 2C). Reversion of U-to-C hyper-edits from all sequences reduced US survivor AAVS and SAVS evolutionary rates to a level that was similar to acute-infection rate estimates (Figures 2A and 2B). While US EVD survivors received multiple therapeutic EVD treatments during early disease, we did not observe any mutations within viral regions (GP, VP35, L) targeted by these compounds (Supplemental Experimental Procedures). Thus, we hypothesize that these *de novo* U-to-C hyper-edits are not the result of therapeutic EVD treatments. U-to-C hyper-editing was also observed in Sierra Leone survivors 1 and 5, but from the available specimens, we cannot determine whether these changes occurred *de novo* during viral persistence, or during acute infection, because other AAVS from Sierra Leone (SLE) share the same set of mutations (Figure 2C).

U-to-C hyper-editing is not unique to SAVS, similar patterns have also been observed within AAVS (Dudas et al., 2017; Ni et al., 2016; Park et al., 2015; Smits et al., 2015; Tong et al., 2015), however, it is currently unknown whether acute- and persistence-specific hyper-edited genomic regions exist. Here, we observed that most acute editing occurred within non-coding regions and the highest rates of hyper-editing were on the

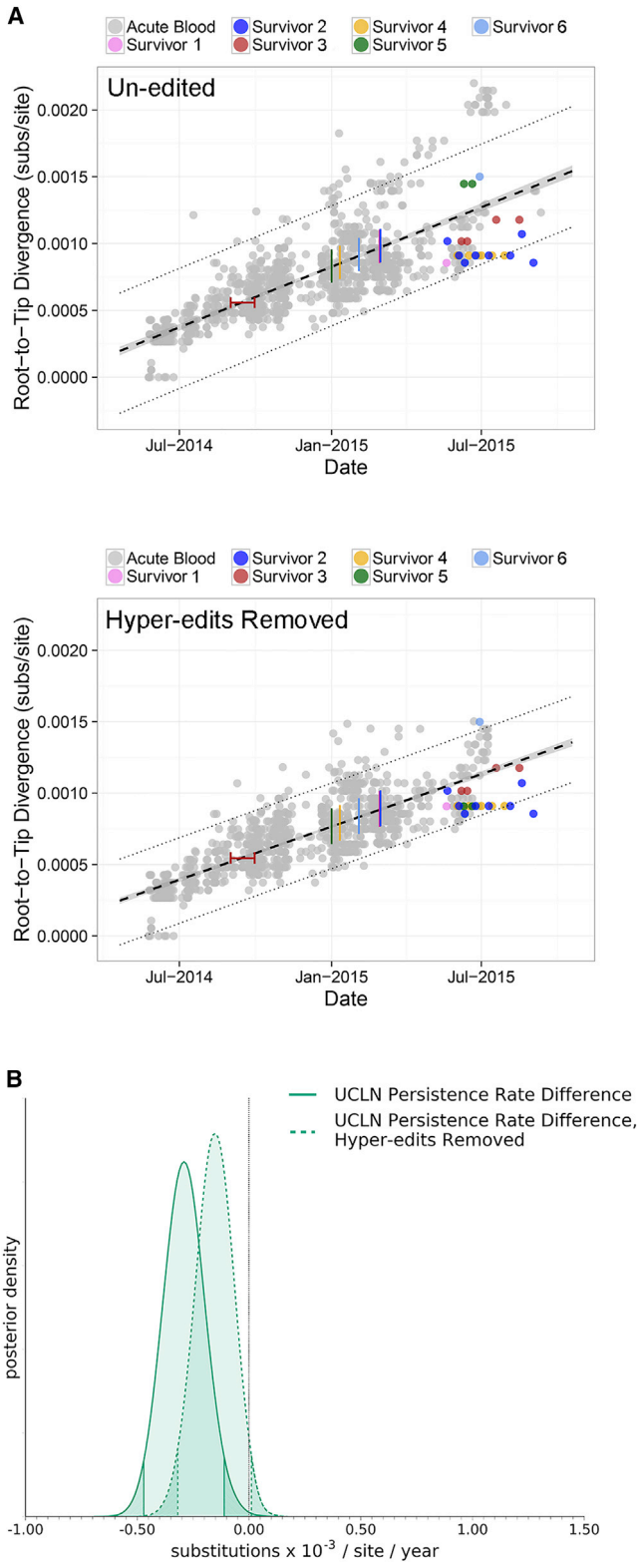


Figure 1. EBOV in Semen Specimens from Sierra Leonean EVD Survivors Exhibit Reduced Evolutionary Rates

(A) Genetic divergence versus specimen collection date for nearly all SLE viral sequences (n = 1,058) acquired from blood, plasma, or oral swab during acute infection (gray) and from semen during persistent infection (color). Colored bars represent survivor-reported symptom onset dates, and red whiskers represent onset date ambiguity for survivor 3. Top: includes sequences without editing. Bottom: includes sequences with reversion of potential U-to-C hyper-edited sites. Acute specimen average divergence from root is black dashed line and corresponding 95% confidence interval is gray (along black dashed line). Dotted lines represent 95% prediction intervals. EVD survivors 1, 2, 3, and 4 exhibited a reduced number of substitutions relative to the mean AAVS divergence, whereas survivors 5 and 6 exhibited an increased number of substitutions relative to the mean AAVS divergence (upper panel). Removal of hyper-edited sites reduced the number of substitutions for patient 5 (bottom).

(B) SAVS exhibit significantly reduced evolutionary rates compared to AAVS. Posterior rate distribution differences of SAVS compared to AAVS using unedited sequences (solid line) and reversion of potential hyper-edited sites (dashed line). Shaded density tails indicate 95% highest posterior density interval (HPD) and black dotted line indicates the expectation that rate estimates are identical during acute and persistent infection.

3′ untranslated NP and VP40 transcripts. (Figure 2C). The distribution of hyper-edited sites in Figure 2C represents a combination of both *de novo* and ancestrally acquired hyper-edits. In contrast to AAVS, edited sites in SAVS are only within a distinct region on the 3′ untranslated NP transcript (Figure 2C). Hypermutation within this region was also observed with high frequency within AAVS and is near a U-to-C editing site (3008-11) that can upregulate NP transcription (Figure 2C) (Ni et al., 2016). Because ADAR editing deaminates adenosine to inosine, which base pairs with cytidine, canonical ADAR editing typically results in A → G mutations on the affected strand (Bass, 2002). Therefore, these U-to-C hyper-edits likely reflect ADAR editing of the viral (RNA−) genome.

Selective Pressures within Immune-Privileged Sites

Because immune-privileged sites represent a unique niche, EBOV may experience selective pressure differences during acute and persistent infection. Selective pressures during acute infection were first estimated using Bayesian robust counting and compared to phylogenetic analysis by maximum likelihood (PAML) branch- and branch-site-specific models. To prevent rate overestimation by double-counting shared amino acids, the glycoprotein was split at the transcriptional editing site into N-terminal (NGP), C-terminal full-length (GP1 carboxy-terminus and GP2, CGP), and secreted GP (SGP_c), and rates were estimated independently for each protein fragment (Figure 4). Inferred selective pressures were similar when estimated using Bayesian robust counting (AAVS only) and paml modeling (AAVS and SAVS) (Figure 3). In general, ω estimates were similar to or reduced compared to previous robust counting estimates (Park et al., 2015; Tong et al., 2015), consistent with purifying selection acting over a longer time period (Figure 3A). A comparison of the changes accumulated in the cohort, including its analysis in the context of the larger outbreak, did not reveal significant differences among groups (Figure S2; Table S2).

Using the branch model, a moderate increase in ω was observed for the carboxy-terminal secreted glycoprotein tail (sGP_c) ($p = 6.13 \times 10^{-5}$) of SAVS (Figure 3B; Table S2). This data were supported by the branch-site model, which provided evidence of site-based positive selection in SAVS occurring at

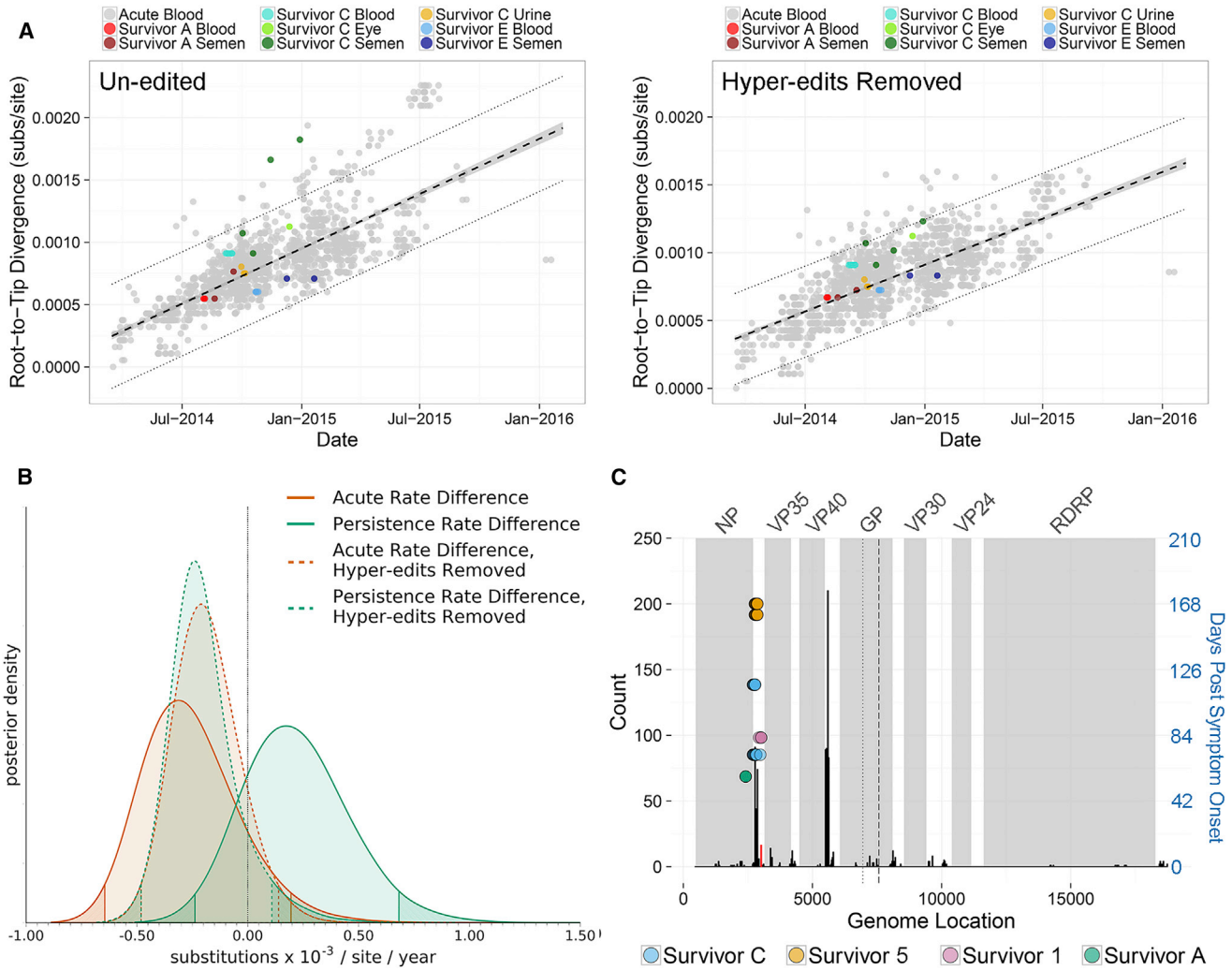


Figure 2. EBOV Sequenced from Acute and Persistent Clinical Specimens Acquired from US EVD Survivors Exhibits Acute-like Evolutionary Rates

(A) Genetic divergence versus specimen collection date for viral sequences from US EVD survivors and 1,498 sequences from SLE, Guinea (GIN), and Liberia (LBR). Left: includes sequences without editing. Right: includes sequences with reversion of potential hyper-edited sites. Viral sequences were acquired from blood, plasma, or oral swab specimens during acute infection (gray), or from blood, plasma, semen, urine, or eye during acute and persistent infection in EVD survivors (color). Mean divergence, 95% confidence interval, and 95% prediction intervals as in Figure 1.

(B) Prior to removal of hyper-edited U-to-C sites, SAVS (green solid line) exhibit ~1.45-fold increased evolutionary rate compared to AAVS (orange solid line). After reversion of U-to-C hyper-edits, SAVS (green dashed line) exhibit a similar divergence as AAVS (orange dashed line). Overall, AAVS and SAVS evolutionary rates were not significantly different from the overall acute evolutionary rate (black dotted line, estimated from AAVS collected in SLE, GIN, and LBR). HPD intervals and rate distribution difference as in Figure 1.

(C) Distribution of U-to-C hyper editing sites using 1,498 sequences from SLE, GIN, and LBR. Occurrence of hyper-editing across the viral genome (black bars) and within coding regions (gray shading). GP transcriptional editing is dotted line, and GP1 and GP2 cleavage is dashed line. Hyper-edited sites from EVD survivors versus days post symptom onset is right y axis (blue). These sites only occurred within a distinct region near the untranslated 3′ nucleoprotein (NP) transcript, which was also observed with high frequency within acute specimens and is near a U-to-C editing site described in Ni et al. (2016) (red bar).

glycoprotein amino acid residues 296N (CGP) (posterior probability 99.9%), 296T (sGP_c) (posterior probability 99.9%), and 315P (sGP_c) (posterior probability 78.2%) (Figure 3C). However, these mutations were each detected in only one EVD survivor (survivor 2: 296N/T and survivor 4: 315P), and thus likely represents an overestimation of ω in SAVS. Therefore, we hypothesize that nonsynonymous changes in sGP_c from SAVS (Figure 3B) are suggestive of the relaxation of selection constraints, rather than evidence of positive selection at specific sites.

Additional unique glycoprotein mutations were observed during viral persistence that were not accurately captured by the PAML analysis. SAVS from survivor 2 contained an insertion in the GP transcriptional editing site (A → AC, 296N/T above)

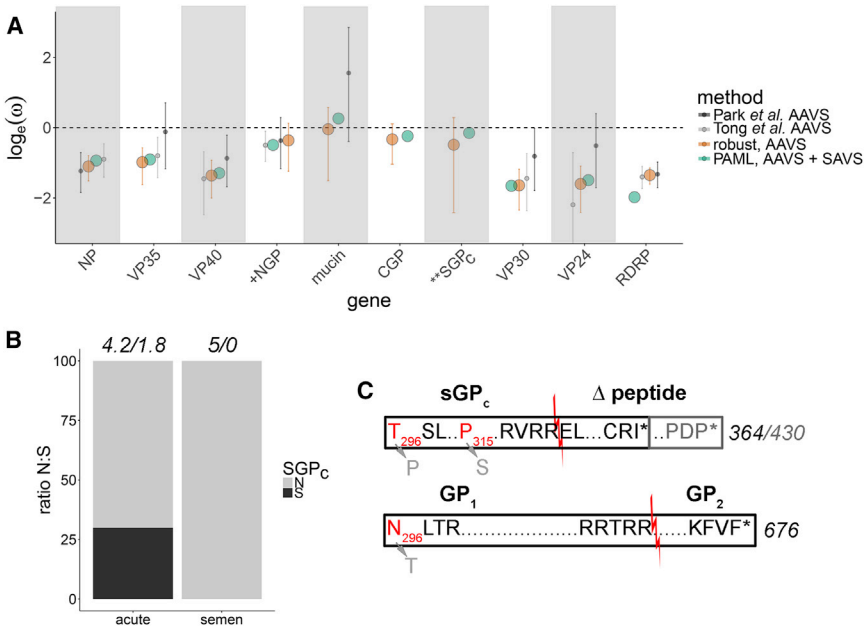


Figure 3. Selective Pressures within the MGT

(A) Comparison of $\log_e(\omega)$ estimates for viral genes calculated using PAML branch model (green) and coalescent robust counting (orange, error bars indicate 95% HPD) or from Park et al. (2015) (dark gray, error bars indicate 95% HPD) and from Tong et al. (2015) (light gray, error bars indicate 95% HPD). Rate estimates in PAML/codeml used SAVS and a subset of AAVS from SLE, GIN, and LBR (collected between 03/2014–09/2015). Robust counting estimates used a subset of AAVS from SLE, GIN and LBR collected between 03/2014–07/2015. Rate estimates from Park et al. (2015) and Tong et al. (2015) were calculated using robust counting with specimens collected between 03/2014–03/2015 and 03/2014–11/2014. In most cases, ω estimates closely agree and were reduced compared to previous estimates, consistent with purifying selection acting over a longer time period. Branch and branch-site PAML models support elevated ω in the secreted GP carboxy-tail from SAVS ("SGP$_c$") (stars). GP rate estimates from Park et al. (2015) and Tong et al. (2015) include full-length GP, rather than partitioned GP, as analyzed here (+ sign).

(B) Comparison of the proportion of total nonsynonymous (N, gray) and synonymous (S, black) counts across AAVS (from SLE, GIN, and LBR) and SAVS tree branches for the SGP$_c$ tail. Numbers above bars are the total count of N/S substitutions summed across AAVS and SAVS branches. Only nonsynonymous substitutions were observed in SAVS within the SGP$_c$ tail.

(C) Comparison of the glycoprotein (GP) C-terminal variants produced following transcriptional RNA editing. Sites identified with the PAML branch-site model to experience potential positive selection in SAVS are in gray and wild-type alleles are in red. Intervening amino acids (not to scale) are summarized with "……" Protease cleavage in the sGP$_c$ tail produces canonical sGP$_c$ and Δ peptide (red line) and cleavage of the full-length GP produces GP1 and GP2. Loss of the sGP stop codon is predicted to produce an extended Δ peptide for survivor 3 (gray).

that shifts the reading frame and results in a viral genome encoding for the full-length GP tail, rather than the canonical sGP tail (Figures 3C and S2A site 6924; Table S2). This insertion was present in all 7 semen specimens from this patient, but its frequency in the SAVS population varied (34%–65%, Figures S2A–S2C). This insertion was also maintained in viral isolates (EBOV grown in tissue culture cells inoculated with survivor 2′s semen specimens) (data not shown), suggesting that it represents a true genomic mutation and not an overrepresentation of edited mRNA in consensus genomes. Interestingly, the end result of this change resembles the 7U/8U mutation that is induced by passage of some strains of EBOV (Zaire, Sudan) in Vero cell lines (Alfson et al., 2015; Volchkova et al., 2011). Additionally, survivor 3 contained a SNP that resulted in the loss of the sGP stop codon, which extends the sGP tail by an additional 66 amino acids (Figure 3C).

Evidence of Active Viral Replication within Semen Specimens from EVD Survivors

Currently there is limited data as to the extent of active viral replication during EBOV persistence and whether this replication occurs with intact or defective viral genomes. Through the use of stranded sequencing and qRT-PCR approaches, we were able to further define the strandedness of viral nucleic acids produced during acute and persistent infection (Figures 4, S3, and S4). Several studies provide support for chronic viral infection occurring due to the production of defective viral genomes (DVGs) containing internal/copy-back deletions (Calain et al., 1999; Li et al., 2011; Tapia et al., 2013) or terminal deletions (Meyer and Schmaljohn, 2000; Meyer and Southern, 1997). Overall, we observed similar depths of negative-sense (i.e., genomic) genome coverage between AAVS and SAVS (Figure 4A). Therefore, we do not see evidence for a preponderance of truncated genomes. However, we did observe a small proportion of chimeric reads containing deletions, duplications or copy back mutations (Table S3). Altogether, we did not observe any consistent trends in the proportion of chimeric reads per patient over time or during acute and persistent infection.

During acute infection, the proportion of positive-sense viral reads varied between 7%–23% (average ± SD: 16.5% ± 6.9%) and during persistent infection between 7%–46% (average 16.0% ± 10.9%) (Figure 4B). As a control, during *in vitro* infection with EBOV-ZsGreen we observed 78%–91% positive-sense viral reads in the monolayer (compatible with the detection of primarily mRNA) and 2%–5% positive-sense viral reads in the supernatant (compatible with the detection of primarily genomic RNA) at 18-48 hr post-infection (Figure S3A). Because SAVS contained proportions of positive-sense reads similar to or greater than that observed during acute infection, these data demonstrate the presence of active transcription/replication in all persistent survivor specimens studied herein.

During acute and persistent infections, the proportion of positive sense reads changed over time. As expected for acute infection, there was an increase in the proportion of positive sense reads over time, consistent with an increase in active viral replication/transcription during EVD (Figure 4C). After recovery

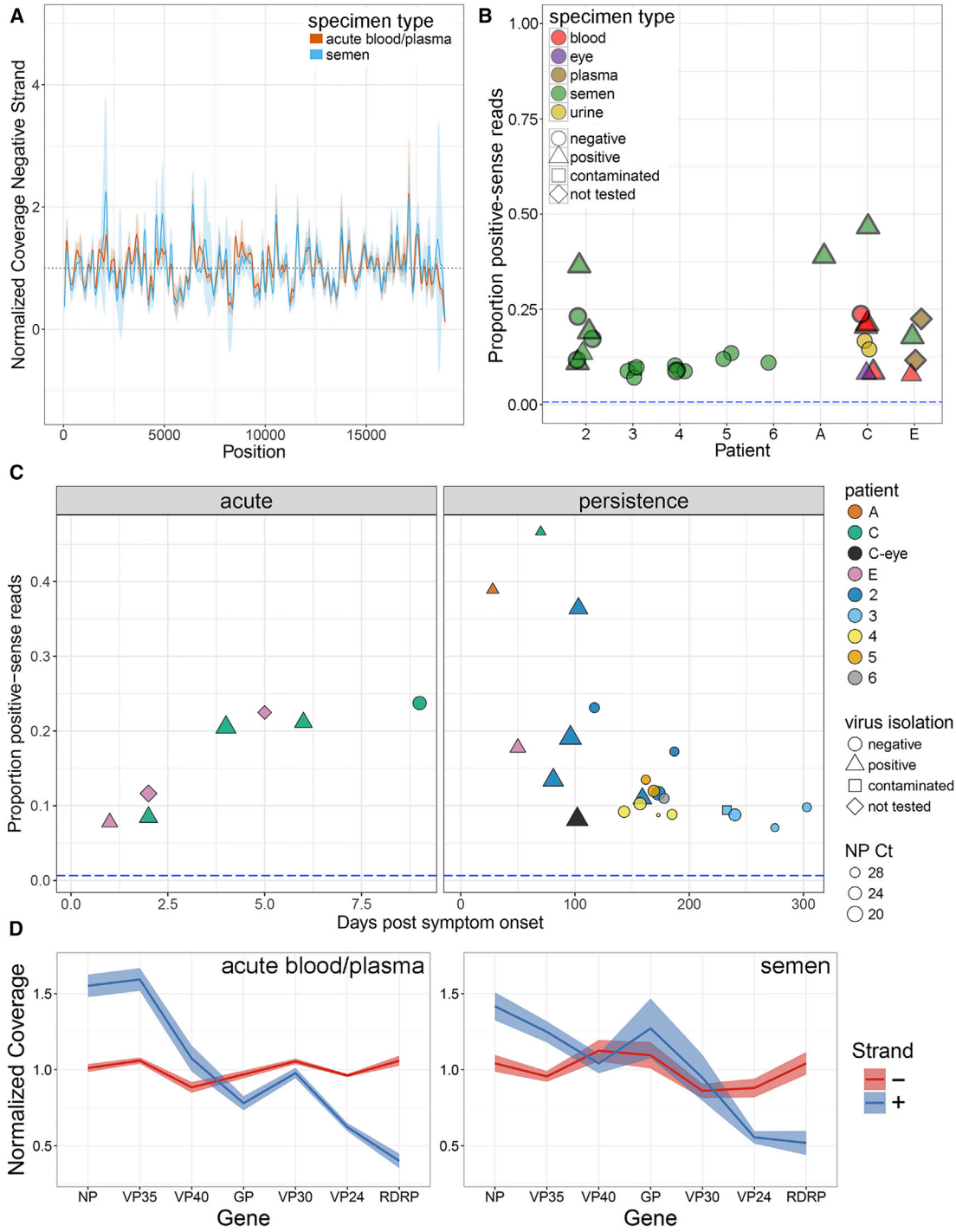


Figure 4. Active Viral Replication during Persistent Infection

(A) Average normalized negative-sense (viral genome) coverage for AAVS and SAVS (coverage mean [line] and standard deviations [shading]).

(B) Proportion of EBOV genome-wide positive-sense reads out of total reads from EVD survivor specimens. Specimen types indicated by color, point shape indicates virus isolation results and specimens in (D) contain thick borders. Blue dashed horizontal line indicates the proportion of positive-sense reads observed from a negative-sense viral RNA *in vitro* transcript (Figure S3A).

(C) Proportion of positive-sense reads versus day post symptom onset for acute specimens (left) and persistent specimens (right). Patients highlighted by color, virus isolation results highlighted by shape and nucleoprotein cycle threshold values highlighted by size.

(D) Proportion of normalized strand-specific reads per EBOV gene from AAVS (left) or SAVS (right). Negative-sense (viral genome) reads in red, and positive-sense (mRNA and viral complementary genome) reads in blue (shading is SE of the normalized coverage means).

from EVD, the ratio of positive-sense reads generally decreased logarithmically with time post onset (Figure 4C), however, in some instances, the proportion of positive-sense reads was higher during persistence than during acute infection (survivors A, C, and 2)—consistent with NP expression from a single survivor (Barnes et al., 2017). For a subset of clinical specimens, we isolated live virus (Spengler et al., 2015; Uyeki et al., 2016; U.S., unpublished data) and observed that the likelihood of positive virus isolation decreased over time (Figures 4B and 4C) and was significantly associated ($p < 0.1$) with the proportion of positive-sense reads (Figures S3B and S3C) and NP Ct value ($p < 0.03$) (Figures S3D and S3E).

We also observed that the relative depth of positive- and negative-sense RNA coverage was consistent with the accepted model of replication for viruses of the order *Mononegavirales*. For AAVS, SAVS, and during an *in vitro* infection, we observed a decrease in positive-sense coverage along the viral genome, consistent with mRNA expression decreasing in a roughly linear manner from the 5′ to 3′ end (Figures 4D, S3F, and S3G). In contrast, there was a steady depth of negative-sense reads across the genome for all specimens, consistent with this strand being synthesized as a continuous RNA molecule (Figures 4D, S3F, and S3G). Similar positive- and negative-sense RNA expression patterns were observed for *in vitro* infected cells (Figures S3F–S3G). However, a slight increase in 5′ negative-sense read coverage was observed during *in vitro* infection (Figures S3F–S3G), and we hypothesize that is due to interrupted/partial negative strand synthesis during active replication.

DISCUSSION

Genomic analysis of EBOV sequences collected from acutely infected and convalescent survivors has yielded important insights into viral replication and selective pressures experienced during acute and persistent infections. During convalescence, EBOV evolutionary rates in the semen, aqueous humor, and urine were either similar to or reduced relative to the rates observed during acute infection in blood and plasma. During persistence, active EBOV replication/transcription continued, but decreased with time, consistent with viral persistence (i.e., long-term viral genome maintenance with active transcription/replication) rather than viral latency (i.e., long-term viral genome maintenance without active replication and low/no transcription). Furthermore, viral persistence did not appear to be linked to defective interfering particles with consistently truncated genomes attenuating wild-type infection (Calain et al., 1999; Li et al., 2011; Meyer and Schmaljohn, 2000; Meyer and Southern, 1997; Tapia et al., 2013). We did observe evidence for a minor population of chimeric reads in both acute and persistent specimens, however, from these short read data, we were not able to estimate the proportion of DVGs in the population, and it is currently unclear what role, if any, these DVGs may play during viral persistence. Finally, EBOV does not appear to have experienced substantially different selective pressures during persistence within immune-privileged niches (testes, eye) as compared to those experienced during acute infections. However, we did observe a moderate relaxation of selective constraints within the sGP carboxy-terminal tail during persistence.

The dichotomy of evolutionary rates observed between the Sierra Leone and US clinical specimens is of particular interest. After reversion of U-to-C hyper-edited sites, Sierra Leonean specimens, on average, exhibited a reduced evolutionary rate, while US specimens exhibited an ''acute-like'' rate. Our observation that SAVS can exhibit a slowed evolutionary rate is in line with a previous rate estimate from a single SAVS (Diallo et al., 2016) and supports rate estimates obtained from sexual transmission cases (Blackley et al., 2016). However, the observation of an ''acute-like'' evolutionary rate during EBOV persistence is a novel finding.

In general, substitution rates represent a complex product of effective population size, mutation rate, generation time, and viral fitness (Duffy et al., 2008). The US and Sierra Leone rate differences are likely due to differences in semen collection times post disease onset; US semen specimens were collected an average of 61 (minimum [min]: 28, maximum [max]: 116) days post onset, whereas Sierra Leonean semen specimens were collected an average of 188 (min: 80, max: 321) days post onset. An acute-like evolutionary rate reflects active ongoing viral replication during early convalescence, whereas the reduced rate may indicate increased pruning of deleterious alleles by purifying selection over time. However, additional factors such as a lower population size, reduced mutation rate, increased generation time, or reduced viral fitness could also contribute to a reduced substitution rate. Because the proportion of positive-sense reads decreases during convalescence, these rate differences also reflect a corresponding decrease in active viral replication over time. Within the MGT, active viral replication could be reduced by the lowered temperatures of the testes, a replication restriction, sequestration of viral nucleic acids into a cellular compartment, and/or immune/apoptotic-mediated clearance. Together, these factors will decrease the viral population size and increase generation time. While immune-privileged sites represent a novel niche, viral fitness differences likely do not contribute to the observed evolutionary rate differences, because we did not observe evidence for significant selective pressure differences in coding regions between SAVS and AAVS.

Viral nucleic acids during acute infection have been detected within the MGT (Dejucq and Jégou, 2001) and many viruses can establish persistent infections within a range of host sites (Randall and Griffin, 2017), however, despite this prevalence, relatively little is known regarding viral evolution during the acute to persistent transition. Previous studies comparing HIV sequences collected from paired blood/PBMCs or semen contained evidence of either compartmentalization or exchange between these two compartments in individual donors (Delwart et al., 1998; Gupta et al., 2000) and those patients that exhibited HIV compartmentalization also exhibited reduced genetic diversity (Pillai et al., 2005). However, abnormally low evolutionary rates for HIV and other viruses (HTLV-I, HTLV-II, SFV, GBV-C, and some plant viruses) are commonly due to a latent viral infection, or slow clonal expansion following viral integration (Duffy et al., 2008)—viral replication strategies that are distinct from models of EBOV replication.

Here, we found that two US survivors (C and A) exhibited evidence of *de novo* U-to-C hyper-editing in specimens acquired

during viral persistence, which inflated the apparent viral evolutionary rate and likely occurred due to host-mediated ADAR1 cytoplasmic editing. Similar excessive ADAR-mediated edits within short regions were also observed within noncoding regions of AAVS (Dudas et al., 2017; Ni et al., 2016; Park et al., 2015; Tong et al., 2015), however, additional molecular studies are needed to confirm that these hyper-edits are due to host enzymes, and/or occur at sites containing secondary structure, and to evaluate the significance of these edits. Preliminary evidence suggests that a U-to-C editing site (3008-11) near those observed within SAVS can upregulate NP transcription (Ni et al., 2016). In other models, loss of ADAR1 editing activity can upregulate interferon-stimulated genes (Rice et al., 2012), thus ADAR-editing of viral transcripts may represent a proviral method to control protein production (hepatitis delta virus), or enhance viral replication (HIV), or may act through an anti-viral method to introduce excessive mutations (LCMV) (Gélinas et al., 2011; Zahn et al., 2007). Similar hyper-editing has also been observed during *in vitro* replication for other viruses (influenza, measles, respiratory syncytial, Epstein-Barr, and polyomavirus) (Iizasa et al., 2010; Kumar and Carmichael, 1997; Martínez and Melero, 2002; Suspène et al., 2011). Most strikingly, U-to-C and G-to-A hyper-editing has been observed following persistent measles infections in the brain 4 and 6 months after initial disease (Baczko et al., 1993; Cattaneo et al., 1988), and a similar pattern of U-to-C edits were observed on the NP 3′ untranslated region during *in vitro* Marburg infection (Shabman et al., 2014). Viral genomes with hyperedits in the VP40 5′ (viral genome orientation) tail were observed in the Magazine Wharf area of SLE after a disease-free 2-week period, potentially representing re-emergence from an EVD survivor, although both of these cases were also associated with ''multiple high-risk contacts'' (Smits et al., 2015; WHO, 2015). While there are some established links between ADAR and interferon signaling (George and Samuel, 1999; Pfaller et al., 2011; Rice et al., 2012), teasing apart the pro- and anti-viral interactions, along with their relationship to viral persistence, will be an important area for future research.

Besides on-the-ground contact tracing, there are currently no molecular signatures that would allow one to confirm whether EBOV was transmitted through contact with an acute case or from contact with an EVD survivor. Here, we observed that a delayed evolutionary rate (as suggested previously by Blackley et al. [2016] and Diallo et al. [2016]) or U-to-C hyper-editing in serial specimens could suggest transmission from persistently infected EVD survivors. However, the absence of these molecular markers does not eliminate persistently infected EVD survivors as potential sources of viral transmission.

Altogether, our data illustrate that EBOV persistence in semen and aqueous humor does not imply a quiescent or latent infection, but instead is an ongoing balance between natural selection and genetic drift within a novel intra-host niche. EBOV persistence within EVD survivors may act as a viral reservoir. Fortunately, sexual transmission of EBOV from EVD survivors is a rare mechanism for viral transmission. Ultimately, understanding the mechanisms of viral persistence in immune-privileged sites will lead to additional treatment options, clarify public health recommendations, and is critical to document whether future or past outbreaks might be due to transmission from persistently infected EVD survivors.

EXPERIMENTAL PROCEDURES

Experimental Model and Subject Details

Human Subjects

Through the joint Sierra Leone Ebola Virus Persistence study (SLEVPS) with the Ministry of Health and Sanitation (MoHS) in Sierra Leone, WHO, China-CDC, and CDC, we had access to semen specimens collected from EVD survivors. The SLEVPS was reviewed and approved by the Sierra Leone Institutional Review Board and the World Health Organization Ethical Review Committee. Acute and persistent specimens from US EVD survivors were collected by their treating physicians and transported to the CDC for detection of viral RNA. This sequencing project was determined by the CDC institutional human subject advisor to be a non-research public health response activity, and institutional review board review was not required.

Method Details

Whole Genome Sequencing and Bioinformatics

RNA was extracted from blood and semen specimens and sequenced using a modified version of the Illumina TruSeq RNA Access Library Prep kit. EBOV genomes were assembled using custom scripts. Additional details are available in the Supplemental Experimental Procedures.

Sequence Analysis

Viral evolutionary rate estimates were conducted using both linear regression modeling and time-structured phylogenies. Additional sequence analysis was conducted using custom-made scripts. Evolutionary selective pressures were estimated using the renaissance counting method in beast/v1.8.2 and hypothesis testing was performed using the codeml model in paml/v4.5. Additional details are available in the Supplemental Experimental Procedures.

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the genomes acquired from clinical specimens reported in this paper are GenBank: KY401638–KY401675 and KY805810-2.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.01.008.

CONSORTIA

Members of the Sierra Leone Ebola Virus Persistence Study Group are: Gibrilla Fadlu Deen, Nathalie Broutet, Barbara Knust, and Wenbo Xu (principal investigators), and James Bangura, Amara Jambai, Faustine James, Alie H. Wurie, Francis Yamba, Foday Sahr, Foday R. Sesay, Thomas A. Massaquoi, Tina Davies, Pierre Formenty, Anna E. Thorsen, Archchun Ariyarajah, Marylin Carino, Antoine Coursier, Kara N. Durski, Ndema Habib, Philippe Gaillard, Sihem Landoulski, Margaret O. Lamunu, Jacyln E. Marrinan, Suzanna L.R. McDonald, Dhamari Naidoo, Neetu Abad, Kyle T. Bernstein, Elizabeth Ervin, John D. Klena, Tasneem Malik, Oliver Morgan, Stuart T. Nichol, Christine Ross, Ute Ströher, Hongtu Liu, William Jun Liu, Yong Xhang, Guizhen Wu, Mifang Liang, and Patricia Ongpin (contributors).

ACKNOWLEDGMENTS

The authors acknowledge William Davis (CDC/NCEZID/VSPB), Amy Schuh (CDC/NCEZID/VSPB), Serena Carroll (CDC/NCEZID), and Mark Whitmer (The NerdWerks, LLC) for their thoughtful review and discussion. The authors acknowledge Mike Purdy (CDC/NCHHSTP), William Fischer (LANL), and Mark Whitmer (The NerdWerks, LLC) for their software and script assistance. Work at US Army Medical Research Institute of Infectious Diseases was funded by the Defense Threat Reduction Agency, project CB10246. Financial support

provided by the WHO, the CDC, the China CDC, the Paul G. Allen Family Foundation, the Sierra Leone Ministry of Health and Sanitation, and the Joint United Nations Program on HIV/AIDS. The WHO acknowledges the financial contribution of the WHO Ebola Response Program, the Paul G. Allen Family Foundation, the UNDP (United Nations Development Program), UNFPA (United Nations Population Fund), UNICEF WHO, World Bank Special Program of Research, Development and Research Training in Human Reproduction (HRP), a cosponsored program executed by the WHO in support of the Sierra Leone Ebola Virus Persistence Study. The authors acknowledge the Emory University Serious Communicable Diseases Program; the Military Hospital 34 and Lungi General Hospital staff; the Sierra Leone Association of Ebola Survivors: Yusuf Kabba; WHO: Ian Askew, Bruce Aylward, Rachel Baggaley, Anshu Banerjee, George Bindi, Rick Brennan, Mauricio Calderon, Ian Crozier, Christopher Dye, Peter Graaff, James N. Kiarie, Marie-Paule Kieny, Anders Nordström, Collins Owili, Lee Sharkey, and Lisa Thomas; and the US-CDC: Dianna Blau (IDP), Aaron Brault (ADB), Jorn Winter (VSDB), Tara Sealy (VSPB), John D. Klena (VSPB), and Lance Presser (VSPB). The content of this publication does not necessarily reflect the views or policies of the US Army. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of CDC.

AUTHOR CONTRIBUTIONS

Conceptualization, S.L.M.W., J.T.L., M.R.W., G.P., and U.S.; Methodology, S.L.M.W., J.T.L., S.S.S., G.P., and U.S.; Software, S.L.W.M., S.S.S., J.T.L., G. Dudas, and A.R.; Validation, S.L.M.W. and J.T.L.; Formal Analysis, S.L.M.W., J.T.L., G. Dudas, A.R., G.P., and U.S.; Investigation, S.L.M.W., J.T.L., M.R.W., K. Patel, K. Prieto, and S.S.S.; Resources, F.S., E.C., B.K., D.N., G. Deen, P.F., S.T.N., G.P., and U.S.; Data Curation, S.L.M.W. and J.T.L.; Writing – Original Draft, S.L.M.W., U.S., and G.P.; Writing – Review & Editing, S.L.M.W., J.T.L., M.R.W., K. Patel, G. Dudas, A.R., B.K., P.F., S.T.N., G.P., and U.S.; Visualization, S.L.M.W., J.T.L., and G. Dudas; Supervision, F.S., B.K., D.N., G. Deen, P.F., S.T.N., G.P., and U.S.; Project Administration, S.L.M.W., B.K., P.F., G.P., and U.S.; and Funding Administration, B.K., P.F., S.T.N., G.P., and U.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 25, 2017
Revised: November 20, 2017
Accepted: January 2, 2018
Published: January 30, 2018

REFERENCES

Alfson, K.J., Avena, L.E., Beadles, M.W., Menzie, H., Patterson, J.L., Carrion, R., Jr., and Griffiths, A. (2015). Genetic changes at the glycoprotein editing site associated with serial passage of Sudan virus. J. Infect. Dis. *212* (*Suppl 2*), S295–S304.

Arias, A., Watson, S.J., Asogun, D., Tobin, E.A., Lu, J., Phan, M.V.T., Jah, U., Wadoum, R.E.G., Meredith, L., Thorne, L., et al. (2016). Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Virus Evol. *2*, vew016.

Baczko, K., Lampe, J., Liebert, U.G., Brinckmann, U., ter Meulen, V., Pardowitz, I., Budka, H., Cosby, S.L., Isserte, S., and Rima, B.K. (1993). Clonal expansion of hypermutated measles virus in a SSPE brain. Virology *197*, 188–195.

Barnes, K.G., Kindrachuk, J., Lin, A.E., Wohl, S., Qu, J., Tostenson, S.D., Dorman, W.R., Busby, M., Siddle, K.J., Luo, C.Y., et al. (2017). Evidence of Ebola virus replication and high concentration in semen of a patient during recovery. Clin. Infect. Dis. *65*, 1400–1403.

Bass, B.L. (2002). RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem. *71*, 817–846.

Bausch, D.G., Towner, J.S., Dowell, S.F., Kaducu, F., Lukwiya, M., Sanchez, A., Nichol, S.T., Ksiazek, T.G., and Rollin, P.E. (2007). Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. J. Infect. Dis. *196* (*Suppl 2*), S142–S147.

Blackley, D.J., Wiley, M.R., Ladner, J.T., Fallah, M., Lo, T., Gilbert, M.L., Gregory, C., D'ambrozio, J., Coulter, S., Mate, S., et al. (2016). Reduced evolutionary rate in reemerged Ebola virus transmission chains. Sci. Adv. *2*, e1600378.

Calain, P., Monroe, M.C., and Nichol, S.T. (1999). Ebola virus defective interfering particles and persistent infection. Virology *262*, 114–128.

Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., ter Meulen, V., and Billeter, M.A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. Cell *55*, 255–265.

Christie, A., Davies-Wayne, G.J., Cordier-Lassalle, T., Blackley, D.J., Laney, A.S., Williams, D.E., Shinde, S.A., Badio, M., Lo, T., Mate, S.E., et al.; Centers for Disease Control and Prevention (CDC) (2015). Possible sexual transmission of Ebola virus - Liberia, 2015. MMWR Morb. Mortal. Wkly. Rep. *64*, 479–481.

Deen, G.F., Knust, B., Broutet, N., Sesay, F.R., Formenty, P., Ross, C., Thorson, A.E., Massaquoi, T.A., Marrinan, J.E., Ervin, E., et al. (2017). Ebola RNA persistence in semen of Ebola virus disease survivors - preliminary report. N. Engl. J. Med. *377*, 1428–1437.

Dejucq, N., and Jégou, B. (2001). Viruses in the mammalian male genital tract and their effects on the reproductive system. Microbiol. Mol. Biol. Rev. *65*, 208–231.

Delwart, E.L., Mullins, J.I., Gupta, P., Learn, G.H., Jr., Holodniy, M., Katzenstein, D., Walker, B.D., and Singh, M.K. (1998). Human immunodeficiency virus type 1 populations in blood and semen. J. Virol. *72*, 617–623.

Diallo, B., Sissoko, D., Loman, N.J., Bah, H.A., Bah, H., Worrell, M.C., Conde, L.S., Sacko, R., Mesfin, S., Loua, A., et al. (2016). Resurgence of Ebola virus disease in Guinea linked to a survivor with virus persistence in seminal fluid for more than 500 days. Clin. Infect. Dis. *63*, 1353–1356.

Dudas, G., Carvalho, L.M., Bedford, T., Tatem, A.J., Baele, G., Faria, N.R., Park, D.J., Ladner, J.T., Arias, A., Asogun, D., et al. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature *544*, 309–315.

Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. *9*, 267–276.

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.

George, C.X., and Samuel, C.E. (1999). Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. Proc. Natl. Acad. Sci. USA *96*, 4621–4626.

Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S., 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.

Gupta, P., Leroux, C., Patterson, B.K., Kingsley, L., Rinaldo, C., Ding, M., Chen, Y., Kulka, K., Buchanan, W., McKeon, B., and Montelaro, R. (2000). Human immunodeficiency virus type 1 shedding pattern in semen correlates with the compartmentalization of viral Quasi species between blood and semen. J. Infect. Dis. *182*, 79–87.

Iizasa, H., Wulff, B.E., Alla, N.R., Maragkakis, M., Megraw, M., Hatzigeorgiou, A., Iwakiri, D., Takada, K., Wiedmer, A., Showe, L., et al. (2010). Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency. J. Biol. Chem. *285*, 33358–33370.

Jacobs, M., Rodger, A., Bell, D.J., Bhagani, S., Cropley, I., Filipe, A., Gifford, R.J., Hopkins, S., Hughes, J., Jabeen, F., et al. (2016). Late Ebola virus relapse causing meningoencephalitis: a case report. Lancet *388*, 498–503.

Kumar, M., and Carmichael, G.G. (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proc. Natl. Acad. Sci. USA *94*, 3542–3547.

Kuming, B.S., and Kokoris, N. (1977). Uveal involvement in Marburg virus disease. Br. J. Ophthalmol. *61*, 265–266.

Ladner, J.T., Wiley, M.R., Mate, S., Dudas, G., Prieto, K., Lovett, S., Nagle, E.R., Beitzel, B., Gilbert, M.L., Fakoli, L., et al. (2015). Evolution and spread of Ebola virus in Liberia, 2014-2015. Cell Host Microbe *18*, 659–669.

Li, D., Lott, W.B., Lowry, K., Jones, A., Thu, H.M., and Aaskov, J. (2011). Defective interfering viral particles in acute dengue infections. PLoS ONE *6*, e19447.

Martínez, I., and Melero, J.A. (2002). A model for the generation of multiple A to G transitions in the human respiratory syncytial virus genome: predicted RNA secondary structures as substrates for adenosine deaminases that act on RNA. J. Gen. Virol. *83*, 1445–1455.

Martini, G.A., and Schmidt, H.A. (1968). [Spermatogenic transmission of the "Marburg virus". (Causes of "Marburg simian disease")]. Klin. Wochenschr. *46*, 398–400.

Mate, S.E., Kugelman, J.R., Nyenswah, T.G., Ladner, J.T., Wiley, M.R., Cordier-Lassalle, T., Christie, A., Schroth, G.P., Gross, S.M., Davies-Wayne, G.J., et al. (2015). Molecular evidence of sexual transmission of Ebola virus. N. Engl. J. Med. *373*, 2448–2454.

Meyer, B.J., and Schmaljohn, C.S. (2000). Persistent hantavirus infections: characteristics and mechanisms. Trends Microbiol. *8*, 61–67.

Meyer, B.J., and Southern, P.J. (1997). A novel type of defective viral genome suggests a unique strategy to establish and maintain persistent lymphocytic choriomeningitis virus infections. J. Virol. *71*, 6757–6764.

Ni, M., Chen, C., Qian, J., Xiao, H.-X., Shi, W.-F., Luo, Y., Wang, H.-Y., Li, Z., Wu, J., Xu, P.-S., et al. (2016). Intra-host dynamics of Ebola virus during 2014. Nat. Microbiol. *1*, 16151.

Park, D.J., Dudas, G., Wohl, S., Goba, A., Whitmer, S.L., Andersen, K.G., Sealfon, R.S., Ladner, J.T., Kugelman, J.R., Matranga, C.B., et al. (2015). Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. Cell *161*, 1516–1526.

Pfaller, C.K., Li, Z., George, C.X., and Samuel, C.E. (2011). Protein kinase PKR and RNA adenosine deaminase ADAR1: new roles for old players as modulators of the interferon response. Curr. Opin. Immunol. *23*, 573–582.

Pillai, S.K., Good, B., Pond, S.K., Wong, J.K., Strain, M.C., Richman, D.D., and Smith, D.M. (2005). Semen-specific genetic characteristics of human immunodeficiency virus type 1 env. J. Virol. *79*, 1734–1742.

Randall, R.E., and Griffin, D.E. (2017). Within host RNA virus persistence: mechanisms and consequences. Curr. Opin. Virol. *23*, 35–42.

Rice, G.I., Kasher, P.R., Forte, G.M., Mannion, N.M., Greenwood, S.M., Szynkiewicz, M., Dickerson, J.E., Bhaskar, S.S., Zampini, M., Briggs, T.A., et al. (2012). Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. Nat. Genet. *44*, 1243–1248.

Rodriguez, L.L., De Roo, A., Guimard, Y., Trappier, S.G., Sanchez, A., Bressler, D., Williams, A.J., Rowe, A.K., Bertolli, J., Khan, A.S., et al. (1999). Persistence and genetic stability of Ebola virus during the outbreak in Kikwit, Democratic Republic of the Congo, 1995. J. Infect. Dis. *179* (*Suppl 1*), S170–S176.

Rowe, A.K., Bertolli, J., Khan, A.S., Mukunu, R., Muyembe-Tamfum, J.J., Bressler, D., Williams, A.J., Peters, C.J., Rodriguez, L., Feldmann, H., et al. (1999). Clinical, virologic, and immunologic follow-up of convalescent Ebola hemorrhagic fever patients and their household contacts, Kikwit, Democratic Republic of the Congo. Commission de Lutte contre les Epidémies à Kikwit. J. Infect. Dis. *179* (*Suppl 1*), S28–S35.

Shabman, R.S., Jabado, O.J., Mire, C.E., Stockwell, T.B., Edwards, M., Mahajan, M., Geisbert, T.W., and Basler, C.F. (2014). Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. MBio *5*, e02011.

Simon-Loriere, E., Faye, O., Faye, O., Koivogui, L., Magassouba, N., Keita, S., Thiberge, J.M., Diancourt, L., Bouchier, C., Vandenbogaert, M., et al. (2015). Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. Nature *524*, 102–104.

Sissoko, D., Duraffour, S., Kerber, R., Kolie, J.S., Beavogui, A.H., Camara, A.M., Colin, G., Rieger, T., Oestereich, L., Pályi, B., et al. (2017a). Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study. Lancet Glob. Health *5*, e80–e88.

Sissoko, D., Keita, M., Diallo, B., Aliabadi, N., Fitter, D.L., Dahl, B.A., Akoi Bore, J., Raymond Koundouno, F., Singethan, K., Meisel, S., et al. (2017b). Ebola virus persistence in breast milk after no reported illness: a likely source of virus transmission from mother to child. Clin. Infect. Dis. *64*, 513–516.

Smits, S.L., Pas, S.D., Reusken, C.B., Haagmans, B.L., Pertile, P., Cancedda, C., Dierberg, K., Wurie, I., Kamara, A., Kargbo, D., et al. (2015). Genotypic anomaly in Ebola virus strains circulating in Magazine Wharf area, Freetown, Sierra Leone, 2015. Euro Surveill. *20*. https://doi.org/10.2807/1560-7917.ES.2015.20.40.30035.

Soka, M.J., Choi, M.J., Baller, A., White, S., Rogers, E., Purpura, L.J., Mahmoud, N., Wasunna, C., Massaquoi, M., Abad, N., et al. (2016). Prevention of sexual transmission of Ebola in Liberia through a national semen testing and counselling programme for survivors: an analysis of Ebola virus RNA results and behavioural data. Lancet Glob. Health *4*, e736–e743.

Spengler, J.R., McElroy, A.K., Harmon, J.R., Ströher, U., Nichol, S.T., and Spiropoulou, C.F. (2015). Relationship Between Ebola Virus Real-Time Quantitative Polymerase Chain Reaction-Based Threshold Cycle Value and Virus Isolation From Human Plasma. J. Infect. Dis. *212* (*Suppl 2*), S346–S349.

Sterk, E. (2008). Filovirus Hemorrhagic Fever Guideline (http://www.slamviweb.org/es/ebola/fhffinal.pdf: Medecins Sans Frontieres).

Suspène, R., Petit, V., Puyraimond-Zemmour, D., Aynaud, M.M., Henry, M., Guétard, D., Rusniok, C., Wain-Hobson, S., and Vartanian, J.P. (2011). Double-stranded RNA adenosine deaminase ADAR-1-induced hypermutated genomes among inactivated seasonal influenza and live attenuated measles virus vaccines. J. Virol. *85*, 2458–2462.

Tapia, K., Kim, W.K., Sun, Y., Mercado-López, X., Dunay, E., Wise, M., Adu, M., and López, C.B. (2013). Defective viral genomes arising in vivo provide critical danger signals for the triggering of lung antiviral immunity. PLoS Pathog. *9*, e1003703.

Tong, Y.G., Shi, W.F., Di, L., Qian, J., Liang, L., Bo, X.C., Liu, J., Ren, H.G., Fan, H., Ni, M., et al. (2015). Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature *524*, 93–96.

Uyeki, T.M., Erickson, B.R., Brown, S., McElroy, A.K., Cannon, D., Gibbons, A., Sealy, T., Kainulainen, M.H., Schuh, A.J., Kraft, C.S., et al. (2016). Ebola virus persistence in semen of male survivors. Clin. Infect. Dis. *62*, 1552–1555.

Varkey, J.B., Shantha, J.G., Crozier, I., Kraft, C.S., Lyon, G.M., Mehta, A.K., Kumar, G., Smith, J.R., Kainulainen, M.H., Whitmer, S., et al. (2015). Persistence of Ebola virus in ocular fluid during convalescence. N. Engl. J. Med. *372*, 2423–2427.

Volchkova, V.A., Dolnik, O., Martinez, M.J., Reynard, O., and Volchkov, V.E. (2011). Genomic RNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. J. Infect. Dis. *204* (*Suppl 3*), S941–S946.

WHO (2014). Clinical Management of Patients with Viral Hemorrhagic Fever: A Pocket Guide for the Front-line Health Worker. https://www.unicef.org/cbsc/files/VHF_pocket_book_Guinea-2014.pdf.

WHO (2015). Ebola Situation Report - 24 June 2015. http://apps.who.int/ebola/current-situation/ebola-situation-report-24-june-2015.

WHO (2016a). Ebola Situation Report - March 30, 2016. http://apps.who.int/ebola/current-situation/ebola-situation-report-30-march-2016.

WHO (2016b). Clinical Care for Survivors of Ebola Virus Disease: Interim Guidance (Geneva, Switzerland: World Health Organization).

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.

Cell Reports, Volume *22*

Supplemental Information

Active Ebola Virus Replication and Heterogeneous Evolutionary Rates in EVD Survivors

Shannon L.M. Whitmer, Jason T. Ladner, Michael R. Wiley, Ketan Patel, Gytis Dudas, Andrew Rambaut, Foday Sahr, Karla Prieto, Samuel S. Shepard, Ellie Carmody, Barbara Knust, Dhamari Naidoo, Gibrilla Deen, Pierre Formenty, Stuart T. Nichol, Gustavo Palacios, Ute Ströher, and Ebola Virus Persistence Study Group

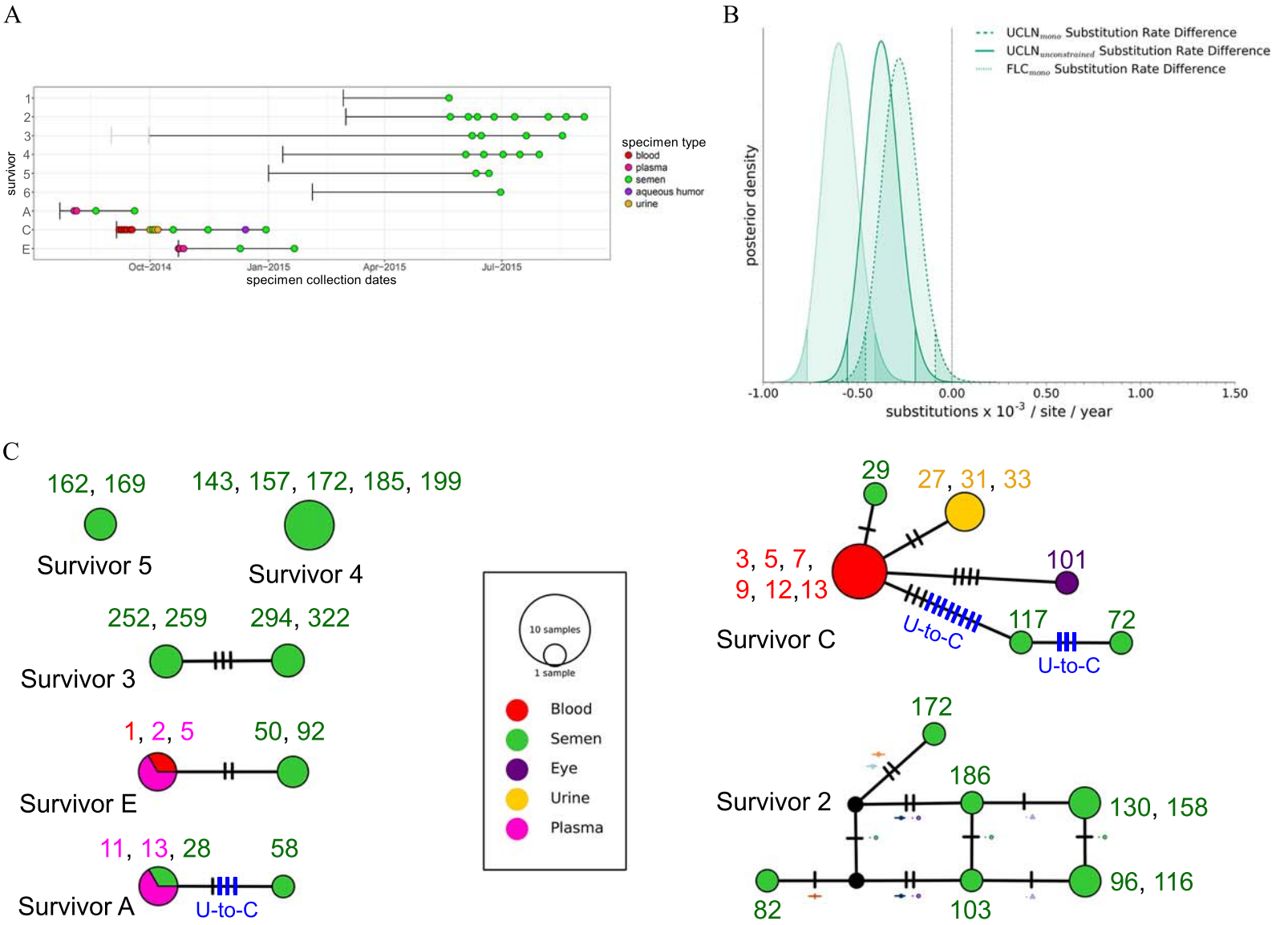


Figure S1: Overview of clinical specimens collected from Ebola virus disease (EVD) survivors, viral evolutionary rates and comparison of viral sequence changes per survivor, Related to Figures 1 and 2. **A)** Overview of clinical specimens collected from Ebola virus disease (EVD) survivors in Sierra Leone (survivors 1, 2, 3, 4, 5, and 6) and in the United States (survivors A, C, E). Survivor-reported symptom onset date is indicated with a black vertical bar, and survivor-reported ambiguity in onset is illustrated with a grey line (survivor 3). Clinical specimens from US EVD survivors were collected during acute and persistent infection, while clinical specimens from Sierra Leonean EVD survivors were collected only during persistent viral infection. Additional specimens were collected from survivors; here we only include specimens that produced a nearly-complete viral genome. **B)** Ebola virus in semen specimens from Sierra Leonean EVD survivors exhibits reduced evolutionary rates. Posterior distribution of evolutionary rate differences from serial semen specimens provided by EVD survivors relative to acute viral evolutionary rates calculated under FLC and UCLN clock models. FLC_{mono} and $UCLN_{mono}$ rates were calculated with SAVS constrained to survivor-specific monophyletic taxons (2, 3, 4, and 5), while $UCLN_{unconstratined}$ rates were calculated without prior assumptions on the tree. Regions within the shaded density tails indicate the 95% highest posterior density interval (HPD) and black dotted line indicates zero rate distribution difference. **C)** Comparison of AAVS and SAVS from EVD survivors. Median joining haplotype networks constructed using AAVS and SAVS from EVD survivors. Vertical bars indicate nucleotide changes (excluding regions that contain N, ? or -, representing less than 1.1% of consensus genomes. A single sequence with low coverage was removed from this figure (KY805812, survivor C)). Nodes are colored according to specimen matrix from which viral sequences were obtained and node size represents the number of clinical specimens. Numbers above nodes represent dayspost symptom onset. For survivor 2, symbols next to vertical bars coincide with iSNVs symbols in Supplemental Figure 2A. SAVS from survivors A (3 sites) and C (11 sites) exhibited potential evidence of human U-to-C hyper-editing following prolonged MGT persistence.

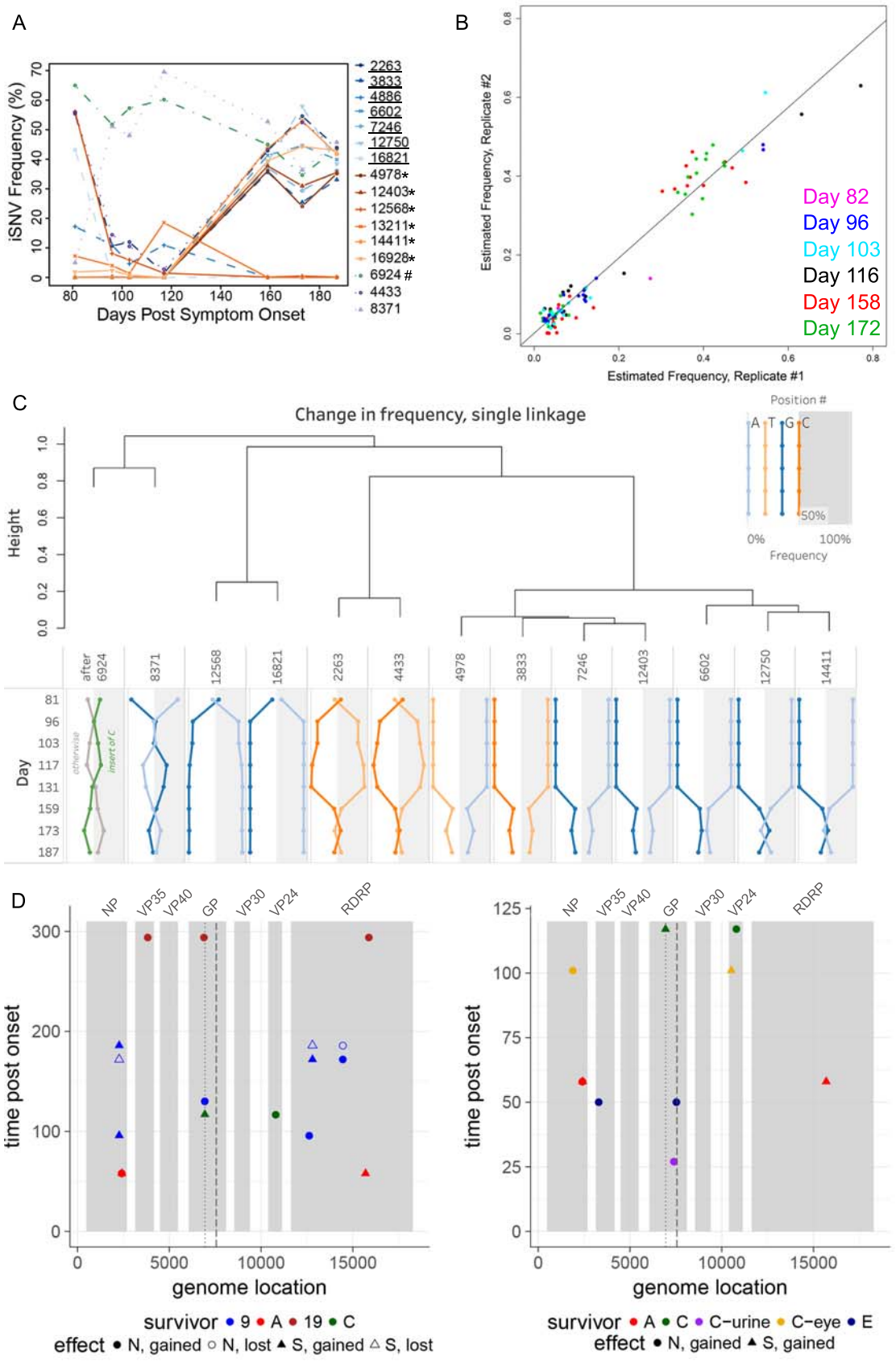


Figure S2: Comparison of AAVS and SAVS from EVD survivors, Related to Figures 2 and 3. **A)** Change in frequency for intrahost single nucleotide variants (iSNVs) (with greater than 15% frequency in a single specimen) versus time post symptom onset for Survivor 2. Sites that result in synonymous (underlined), nonsynonymous (starred), or frameshift (hash) mutations are highlighted and sites without annotations occur in noncoding regions. **B)** Resequencing of technical duplicates yields a similar correlation in iSNV frequencies for SAVS from survivor 2 (r^2=0.9515). **C)** Frequency of intrahost single nucleotide variants (iSNVs) versus time post symptom onset for Survivor 2. A pairwise (Manhattan) distance matrix was computed for each position-allele combination with the vector of the observed frequencies ordered by specimen date. The matrix was used to generate a single-linkage dendrogram (top). Frequency line graphs of iSNV positions, major/minor alleles, and specimen dates were ordered by their position in the dendrogram. Key in upper right-hand corner illustrates allele state (major or minor - grey shading) and value (A,T,C, or G). The presence of co-varying frequency changes suggests either: 1) distinct viral sub-populations, and/or 2) epistasis at the co-varying sites. **D-E)** Acquisition/Loss of synonymous (S) or nonsynonymous (N) changes in SAVS compared to earliest SAVS or AAVS from each survivor. **D)** Coding region changes for SAVS compared to earliest available SAVS from each survivor. Dotted line indicates glycoprotein editing site and dashed line indicates GP1/2 cleavage site. **E)** Coding region changes for SAVS compared to earliest available AAVS from each survivor. Dotted line indicates glycoprotein editing site and dashed line indicates GP1/2 cleavage site.

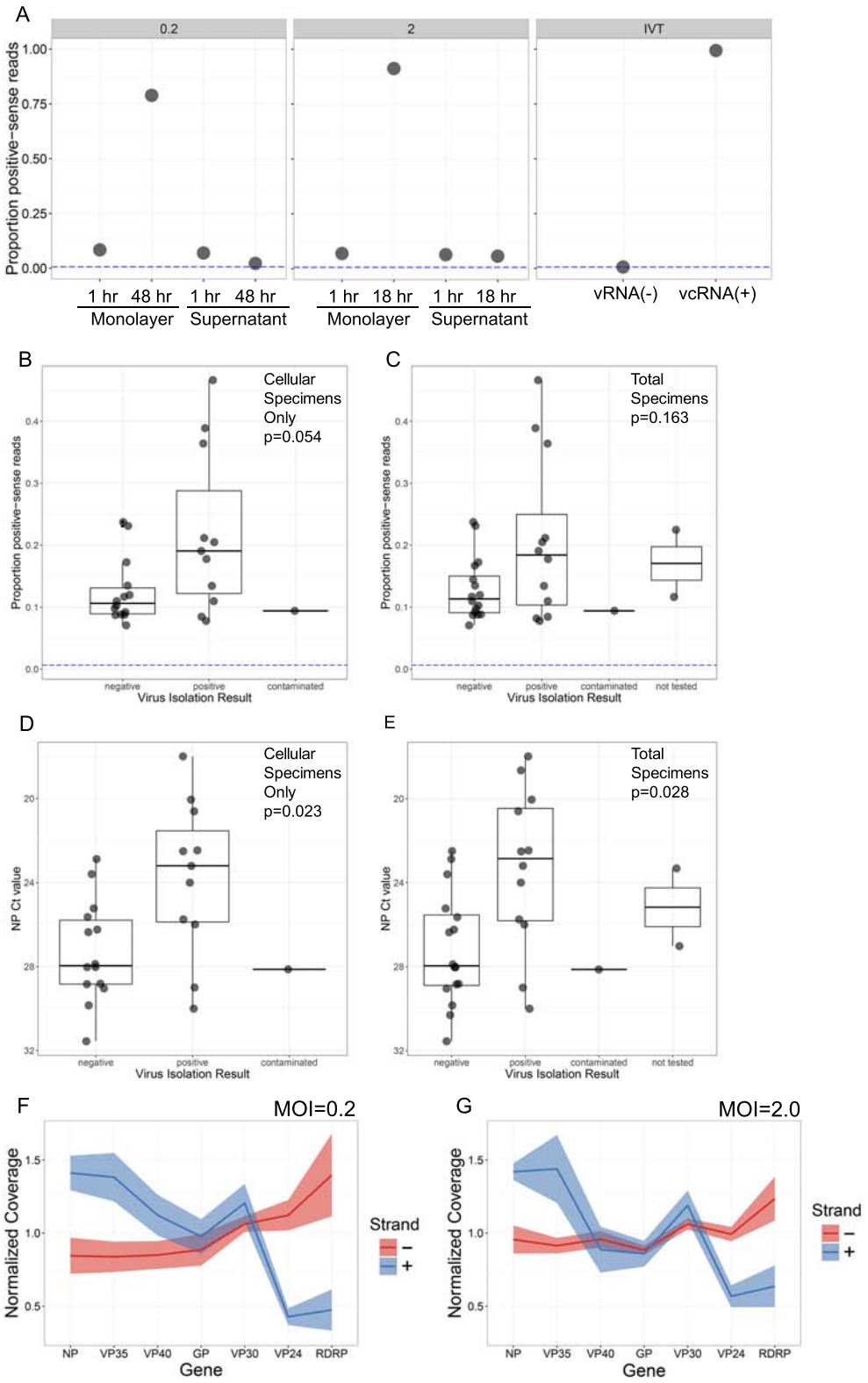


Figure S3: Supportive evidence for active viral replication during persistent infection, Related to Figure 4. To confirm the presence of positive-sense reads from SAVS, we validated our NGS assay with RNA extracted from Huh7 cells infected with recombinant Ebola virus encoding for ZsGreen protein (EBOV-ZsGreen) and *in vitro* transcribed RNA. **A)** Proportion of EBOV genome-wide positive-sense reads sequenced with NGS from an *in vitro* infection of Huh7 cells done at an MOI of 0.2 (1 and 48hpi) (left panel) or 2.0 (1 and 18hpi) (middle panel). Right panel indicates proportion of EBOV genome-wide positive-sense reads from the *in vitro* transcription(IVT) of a negative-sense (vRNA(-)) or positive (vcRNA(+)) viral transcript. **B)** One-sided ANOVA indicates a modest relationship between the proportion of positive-sense reads and virus isolation results ($p=0.054$). This analysis was conducted on clinical specimens containing only cellular material (blood and semen). Maxima and minima in boxplot illustrates the 25th and 75th percentiles, black line indicates median values, whiskers indicate the highest/lowest values within 1.5x the inter-quartile range. **C)** One-sided ANOVA indicates a limited relationship between the proportion of positive-sense reads and virus isolation results ($p=0.163$). This analysis was conducted on clinical specimens containing both acellular (urine, aqueous humor, and plasma) and cellular material (blood and semen). Boxplot values are described in panel B. **D)** One-sided ANOVA indicates a statistically significant ($p<0.05$) relationship between NP real-time polymerase chain reaction (RT-PCR) cycle threshold (Ct) value and virus isolation results ($p=0.023$). This analysis was only conducted on clinical specimens containing cellular material (blood and semen). Boxplot values are described in panel B. **E)** One-sided ANOVA indicates a statistically significant relationship between NP Ct value and virus isolation results ($p=0.028$). This analysis was conducted on clinical specimens containing both acellular (urine, aqueous humor, and plasma) and cellular material (blood and semen). Boxplot values are described in panel B. **F)** Proportion of strand-specific reads per EBOV gene (normalized to total positive- or negative-sense reads) from *in vitro* infection of Huh7 cells at MOI of 0.2. Data represents monolayer and supernatant samples collected after one hour and 48hrs post infection. Negative-sense reads in red, and positive-sense reads in blue. **G)** Proportion of strand-specific reads per EBOV gene (normalized to total positive- or negative-sense reads) from *in vitro* infection of Huh7 cells at MOI of 2. Data represents monolayer and supernatant samples collected after one hour and 18 hrs post infection. Negative-sense reads in red, and positive-sense reads in blue.

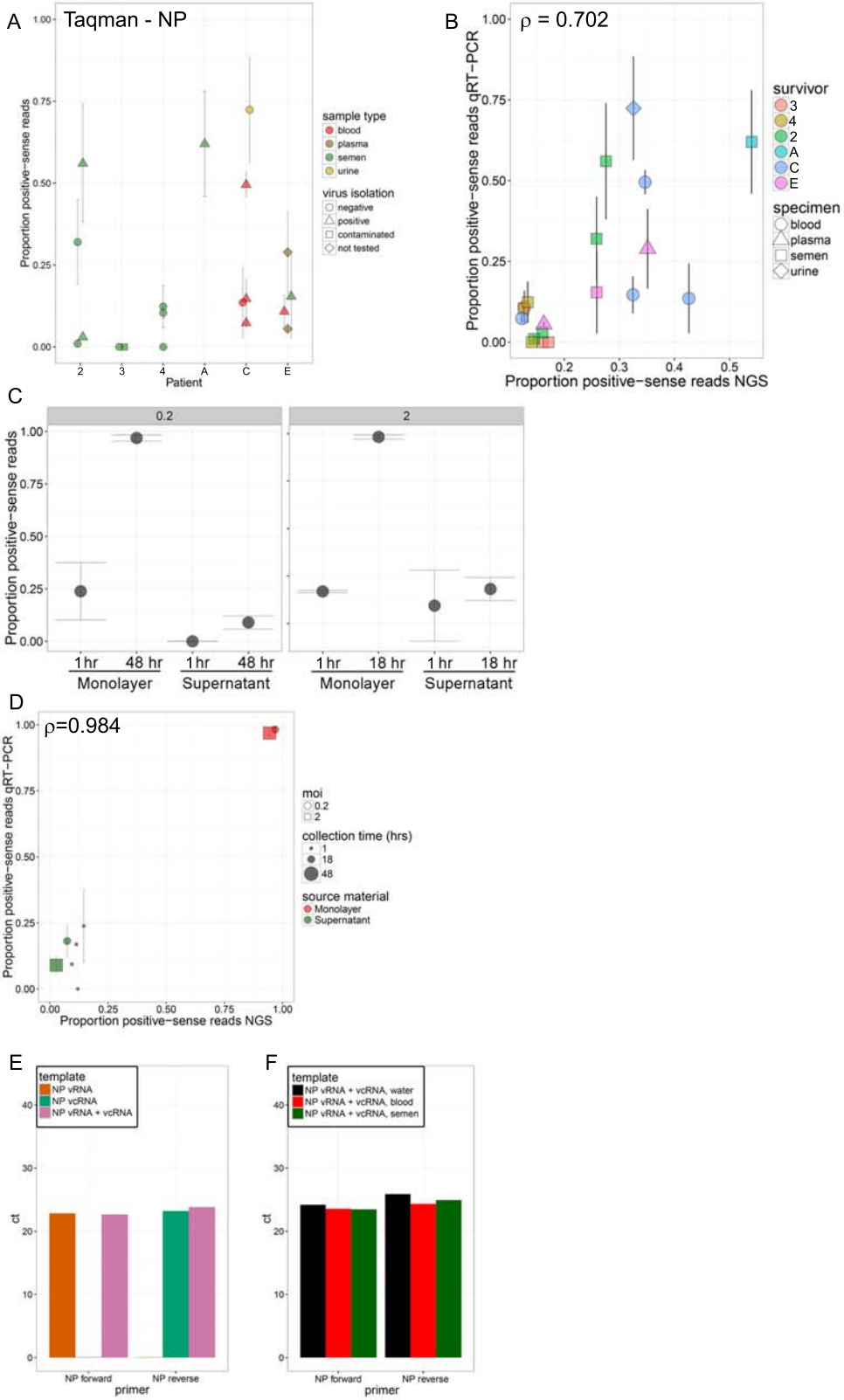


Figure S4: Supportive qRT-PCT evidence of active viral replication during persistent infection, Related to Figure 4. NGS strandedness assay results were further verified by performing strand-specific qRT-PCR with synthetic positive- and negative-sense RNA, RNA remaining from clinical specimens and RNA extracted from Huh7 cells infected with EBOV-ZsGreen. **A)** Proportion of EBOV NP-specific positive-sense reads from each EVD survivor specimen. Specimen types are highlighted with different colors and error bars indicate standard deviations between biological replicates. Virus isolation was attempted on most specimens,and point shape indicates virus isolation results. **B)** Correlation between the proportions of NP-specific positive-sense reads detected by stranded next-generation sequencing and qRT-PCR using RNA extracted from the semen of EVD survivors. Error bars indicate standard deviations in copy numbers detected by qRT-PCR between biological replicates. A positive monotonic relationship was detected as measured by Spearman's rank-order correlation (ρ=0.702). **C)** Proportion of EBOV NP-specific positive-sense reads from an in vitro infection of Huh7 cells done at an MOI of 0.2 (1 and 48hpi) or 2 (1 and 18hpi). Error bars indicate standard deviations in copy numbers detected by qRT-PCR between biological replicates. **D)** Correlation between the proportions of NP-specific positive-sense reads detected by stranded next-generation sequencing and qRT-PCR using RNA extracted from an in vitro infection of Huh7 cells. Error bars indicate standard deviations in copy numbers detected by qRT-PCR between biological replicates. A strongly positive monotonic relationship was detected as measured by Spearman's rank-order correlation (ρ=0.984). **E)** Specificity of NP stranded qRT-PCR assays. Specificity of forward and reverse qRT-PCT assays was assessed and confirmed using negative or positive-sense synthetic RNA, or a mixture of the two strands. **F)** Specificity of NP stranded qRT-PCR assay. Specificity of forward and reverse qRT-PCT assays was assessed and confirmed using negative and positive-sense synthetic RNA. Mixtures of synthetic RNA were spiked into water, or RNA extracted from normal human blood or semen.

Table S1: Evolutionary Rate Estimates from non-edited and edited SAVS from SLE and US EVD Survivors, Related to Figures 1 and 2.

SLE EVD SURVIVORS:

Clock Model	Rate Estimates (*10^3 subs/site/year) mean [95%HPD lower - upper]	
	Acute Rate	Latent Rate
No Edits	0.963 (0.863-1.066)	0.674 (0.504-0.846)
Hyper-edits removed	0.895 (0.805-0.991)	0.739 (0.580-0.904)

Clock Model	Rate Estimates (*10-3 subs/site/year) mean [95%HPD lower - upper]		
	Survivor 1	Survivor 2	Survivor 3
No Edits	0.500 (0.122-0.969)	0.631 (0.418-0.853)	0.862 (0.441-1.339)
Hyper-edits removed	0.631 (0.238-1.068)	0.710 (0.514-0.918)	0.856 (0.529-1.209)

Clock Model	Rate Estimates (*10-3 subs/site/year) mean [95%HPD lower - upper]		
	Survivor 4	Survivor 5	Survivor 6
No Edits	0.717 (0.391-1.071)	0.787 (0.230-1.502)	0.751 (0.298-1.261)
Hyper-edits removed	0.776 (0.498-1.042)	0.815 (0.367-1.334)	0.783 (0.383-1.225)

Clock Model - Uneditted	Loglikelihood Path Sampling		Loglikelihood Stepping Stone	
	Average	Standard Deviation	Average	Standard Deviation
Relaxed Clock (UCLN), Skygrid	*-34238.4*	*7.9*	*-34246.3*	*7.9*
Relaxed Clock (UCLN), constant population	-34359.0	4.4	-34363.1	5.1
Relaxed Clock (UCLN) - monophyly, constant population	-34331.5	4.4	-34340.3	5.1
Fixed Local Clock - monophyly, individual rates, constant population	-34427.0	10.1	-34434.7	6.5
Fixed Local Clock - monophyly, persistent rate, constant population	-34365.1	10.0	-34369.7	9.2

Clock Model - Hyper-edits removed	Loglikelihood Path Sampling		Loglikelihood Stepping Stone	
	Average	Standard Deviation	Average	Standard Deviation
Relaxed Clock (UCLN), Skygrid	*-33822.7*	*4.9*	*-33830.6*	*6.6*
Relaxed Clock (UCLN), constant population	-33939.6	3.7	-33946.7	3.8
Relaxed Clock (UCLN) - monophyly, constant population	-33901.5	7.3	-33912.0	13.6
Fixed Local Clock - monophyly, individual rates, constant population	-33928.6	2.2	-33940.2	4.7
Fixed Local Clock - monophyly, persistent rate, constant population	ND	ND	ND	ND

US EVD SURVIVORS:

Clock Model	Rate Estimates (subs/site/year) mean (95%HPD lower - upper)		
	Acute Rate	Blood Rate	Persistence Rate
No Edits	1.152 (1.043-1.267)	0.884 (0.518-1.290)	1.290 (0.903-1.713)
Hyper-edits removed	1.041 (0.937-1.138)	0.888 (0.577-1.235)	0.859 (0.617-1.133)

Clock Model	Rate Estimates (subs/site/year) mean (95%HPD lower - upper)	
	Survivor A - Acute	Survivor A - Semen
No Edits	1.037 (0.257-2.103)	1.378 (0.426-2.619)
Hyper-edits removed	0.979 (0.331-1.805)	0.816 (0.264-1.499)

Clock Model	Rate Estimates (subs/site/year) mean (95%HPD lower - upper)	
	Survivor E - Acute	Survivor E - Semen
No Edits	0.962 (0.530-1.475)	0.571 (0.132-1.137)
Hyper-edits removed	0.848 (0.361-1.419)	0.634 (0.187-1.174)

Clock Model	Rate Estimates (subs/site/year) mean (95%HPD lower - upper)		
	Survivor C - Acute	Survivor C - Semen and Urine	Survivor C - Eye
No Edits	0.962 (0.530-1.475)	1.431 (0.940-1.956)	0.903 (0.264-1.662)
Hyper-edits removed	0.948 (0.572-1.365)	0.916 (0.619-1.241)	0.930 (0.306-1.669)

Clock Model - uneditted	Loglikelihood Path Sampling		Loglikelihood Stepping Stone	
	Average	Standard Deviation	Average	Standard Deviation
Relaxed Clock (UCLN), Skygrid	*-38288.5*	*9.3*	*-38302.7*	*8.7*
Relaxed Clock (UCLN), constant population	-38456.4	8.8	-38470.6	8.9
Relaxed Clock (UCLN) - monophyly, constant population	-38406.7	5.7	-38420.0	8.9
Fixed Local Clock - monophyly, individual rates, constant population	-38621.5	12.5	-38643.5	24.2
Fixed Local Clock - monophyly, blood/persistent rates, constant population	-38547.7	1.7	-38561.8	9.8

Clock Model - hyper-edits removed	Loglikelihood Path Sampling		Loglikelihood Stepping Stone	
	Average	Standard Deviation	Average	Standard Deviation
Relaxed Clock (UCLN), Skygrid	*-37424.7*	*9.0*	*-37438.1*	*8.7*
Relaxed Clock (UCLN), constant population	-37578.0	8.5	-37591.7	8.5
Relaxed Clock (UCLN) - monophyly, constant population	-37503.8	9.8	-37529.9	16.4
Fixed Local Clock - monophyly, individual rates, constant population	-37680.3	12.0	-37696.0	5.5
Fixed Local Clock - monophyly, blood/persistent rates, constant population	ND	ND	ND	ND

Table S1: Evolutionary Rate Estimates from non-edited and edited SAVS from SLE and US EVD Survivors, Related to Figures 1 and 2.**(TOP)** Bayesian analysis conducted using UCLNunconstrained clock models with un-edited viral sequences and U-to-C hyper-edits removed from viral sequences. Marginal likelihood values from path sampling and stepping stone analysis with different clock models and prior tree assumptions (Relaxed UCLNunconstrained, Relaxed UCLNmonophyletic, Fixed local clockmonophyletic-individual rates, and Fixed local clockmonophyletic-latent rates) are included on lower half. **(BOTTOM)** Evolutionary Rate Estimates from non-edited and edited SAVS using AAVS and SAVS from US EVD Survivors. Bayesian analysis conducted using UCLNunconstrained clock models with un-edited viral sequences and U-to-C hyper-edits removed from viral sequences. Marginal likelihood values from path sampling and stepping stone analysis using different clock models and prior tree assumptions (Relaxed UCLNunconstrained, Relaxed UCLNmonophyletic, Fixed local clockmonophyletic-individual rates, and Fixed local clockmonophyletic-latent rates) are included on lower half.

Table S2: Evolutionary Preassure and iSNV Analysis, Related to Figure 3.

"branch models" (codeml) AAVS vs SAVS

Gene	2ΔlnL	Degrees of Freedom	P value (Bonferroni corrected)
NP	0.67	1	0.4118
VP35	3.04	1	0.0814
VP40	4.29	1	0.0384
NGP	0.01	1	0.9395
Mucin	0.05	1	0.8258
CGP	1.39	1	0.2377
SGP without p19 tail	16.06	1	**6.13E-05**
SGP with p19 tail	5.03	1	0.0249
VP30	0.00	1	0.9517
VP24	2.32	1	0.1276
RDRP	-61.67	1	0.0000

"branch models" (codeml) AAVS vs $SAVS_{fast}$ vs $SAVS_{slow}$

Gene	2ΔlnL	Degrees of Freedom	P value (Bonferroni corrected)
NP	0.73	1	0.3921
VP35	3.05	1	0.0806
VP40	4.01	1	0.0451
NGP	0.15	1	0.7030
Mucin	1.09	1	0.2957
CGP	2.40	1	0.1213
SGP without p19 tail	0.92	1	0.3382
SGP with p19 tail	5.03	1	0.0249
VP30	0.00	1	0.9495
VP24	2.32	1	0.1276
RDRP	-65.01	1	0.0000

"branch-site models" (codeml) AAVS vs SAVS

Gene	2ΔlnL	Degrees of Freedom	P value	Sites under positive selection (NEB)
NP	1.79	1	0.1812	101 E 0.940, 376 Y 0.939
VP35	0.75	1	0.3856	51 P 0.880
VP40	1.67	1	0.1958	131 Q 0.958*, 252 V 0.960*
NGP	5.40E-04	1	0.9815	
Mucin	1.14	1	0.2857	213 P 0.879, 264 T 0.879
CGP	24.91	1	**6.00E-07**	**296 N 0.999****
SGP without p19 tail	24.89	1	**6.06E-07**	**296 T 0.999**, 315 P 0.782**
SGP with p19 tail	24.935648	1	**5.93E-07**	**296 T 0.999**, 315 P 0.766**
VP30	0.01	1	0.9166	
VP24	0.71	1	0.3988	117 R 0.918
RDRP	-3.02E-03	1	0.0000	

Position	Gene	Major Variant	Minor Variant	Effect	Major Amino Acid	Minor Amino Acid
2263	NP	TC**C**	TC**T**	synonymous	S598	S
3833	VP35	TT**T**	TT**C**	synonymous	F235	F
4433	noncoding	C	T	N/A		
4886	VP40	AA**T**	AA**C**	synonymous	N136	N
4978	VP40	C**A**A	C**T**A	nonsynonymous	Q167	L
6602	GP - shared with FL and sGP	CA**A**	CA**G**	synonymous	Q188	Q
6924	GP - shared with FL and sGP	AAA AAA ACC CTC	AAA AAA ACC -TC A	nonsynonymous - results in frame shift from full length GP to sGP	KKTL, full length GP tail	KKTS, sGP tail
7246	full length GP	CA**A**	CA**G**	synonymous	Q108	Q
8371	noncoding	A	G	N/A		
12403	polymerase	**A**TG	**G**TG	nonsynonymous	M275	V
12568	polymerase	**G**CC	**A**CC	nonsynonymous	A330	T
12750	polymerase	AA**A**	AA**G**	synonymous	K390	
13211	polymerase	C**A**A	C**G**A	nonsynonymous	Q544	R
14411	polymerase	G**A**G	G**G**G	nonsynonymous	E944	G
16821	polymerase	TC**A**	TC**G**	synonymous	S1747	S
16928	polymerase	**A**CC	**A**TC	nonsynonymous	T1783	I

Table S2: Evolutionary Preassure and iSNV Analysis, Related to Figure 3. **(TOP)** Likelihood ratio test statistics from PAML branch- and branch-site models. **(BOTTOM)** Effect of iSNV's from SLE Survivor 2 on viral coding and noncoding regions.

Table S3: Chimeric Reads from Sierra Leone and US EVD Survivors, and Cell Culture *in vitro* Infections, Related to Figure 4.

SIERRA LEONE EVD SURVIVORS:

Survivor:	Days post Onset:	Specimen Number:	Chimera Type:	#Unique Deletions:	#Reads Chimeric:	Avg. # Reads Per Chimera:	Standard Deviation Reads Per Chimera	Total # Mapped Reads:	Proportion Mapped Chimeric Reads:
3	252	VP1201500050	Deletions	11	57	5.1818	4.9326	225885	0.0003
3	252	VP1201500050	SmallDups	4	14	3.5	1.5	225885	0.0001
3	252	VP1201500050	LargeDups	8	33	4.125	4.3714	225885	0.0001
3	252	VP1201500050	CopyBacks	1	4	4	0	225885	0
3	259	VP1201500100	Deletions	9	47	5.2222	5.0723	2210508	0
3	259	VP1201500100	LargeDups	13	44	3.3846	2.1318	2210508	0
3	259	VP1201500100	CopyBacks	2	2	1	0	2210508	0
3	294	VP1201500247	Deletions	2	21	10.5	6.5	45989	0.0005
3	294	VP1201500247	SmallDups	2	29	14.5	4.5	45989	0.0006
3	294	VP1201500247	LargeDups	1	2	2	0	45989	0
3	322	VP1201500357	Deletions	2	23	11.5	9.5	24405	0.0009
3	322	VP1201500357	LargeDups	1	4	4	0	24405	0.0002
4	143	VP1201500033	Deletions	3	97	32.3333	28.1227	397245	0.0002
4	143	VP1201500033	SmallDups	1	20	20	0	397245	0.0001
4	143	VP1201500033	LargeDups	7	198	28.2857	27.7731	397245	0.0005
4	143	VP1201500033	CopyBacks	2	54	27	24	397245	0.0001
4	157	VP1201500118	Deletions	1	15	15	0	328497	0
4	157	VP1201500118	SmallDups	6	150	25	21.7486	328497	0.0005
4	157	VP1201500118	LargeDups	3	71	23.6667	30.6522	328497	0.0002
4	172	VP1201500193	Deletions	12	53	4.4167	3.0127	310225	0.0002
4	172	VP1201500193	SmallDups	7	37	5.2857	3.3685	310225	0.0001
4	172	VP1201500193	LargeDups	23	121	5.2609	5.4231	310225	0.0004
4	172	VP1201500193	CopyBacks	3	64	21.3333	18.625	310225	0.0002
4	185	VP1201500235	Deletions	6	10	1.6667	1.1055	5016545	0
4	185	VP1201500235	SmallDups	27	154	5.7037	4.8672	5016545	0
4	185	VP1201500235	LargeDups	28	215	7.6786	10.1526	5016545	0
4	185	VP1201500235	CopyBacks	5	43	8.6	7.3103	5016545	0
4	199	VP1201500293	SmallDups	1	28	28	0	19305	0.0015
5	169	VP1201500132	Deletions	7	35	5	9.396	3363725	0
5	169	VP1201500132	SmallDups	2	43	21.5	1.5	3363725	0
5	169	VP1201500132	LargeDups	3	4	1.3333	0.4714	3363725	0
5	169	VP1201500132	CopyBacks	6	6	1	0	3363725	0
6	178	VP1201500297	SmallDups	1	2	2	0	28892	0.0001
6	178	VP1201500297	LargeDups	2	11	5.5	4.5	28892	0.0004
2	82	VP1201500009	Deletions	2	11	5.5	0.5	204315	0.0001
2	82	VP1201500009	SmallDups	1	11	11	0	204315	0.0001
2	82	VP1201500009	LargeDups	6	59	9.8333	8.1938	204315	0.0003
2	82	VP1201500009	CopyBacks	1	1	1	0	204315	0
2	96	VP1201500046	Deletions	14	76	5.4286	5.0244	8123282	0
2	96	VP1201500046	SmallDups	7	49	7	10.1419	8123282	0
2	96	VP1201500046	LargeDups	20	111	5.55	5.6963	8123282	0
2	96	VP1201500046	CopyBacks	14	81	5.7857	8.6204	8123282	0
2	103	VP1201500084	Deletions	35	748	21.3714	31.5722	1153901	0.0006
2	103	VP1201500084	SmallDups	71	1262	17.7746	28.1085	1153901	0.0011
2	103	VP1201500084	LargeDups	43	682	15.8605	17.2742	1153901	0.0006
2	103	VP1201500084	CopyBacks	72	1208	16.7778	44.6555	1153901	0.001
2	116	VP1201500163	Deletions	7	78	11.1429	7.8272	550976	0.0001
2	116	VP1201500163	SmallDups	6	63	10.5	9.4472	550976	0.0001
2	116	VP1201500163	LargeDups	4	93	23.25	37.963	550976	0.0002
2	116	VP1201500163	CopyBacks	1	3	3	0	550976	0
2	158	VP1201500320	Deletions	38	80	2.1053	1.5181	5205496	0
2	158	VP1201500320	SmallDups	29	154	5.3103	8.9022	5205496	0
2	158	VP1201500320	LargeDups	43	379	8.814	37.1911	5205496	0.0001
2	158	VP1201500320	CopyBacks	14	399	28.5	70.5111	5205496	0.0001
2	172	VP1201500374	Deletions	12	319	26.5833	56.1508	5339682	0.0001
2	172	VP1201500374	SmallDups	16	1187	74.1875	90.4898	5339682	0.0002
2	172	VP1201500374	LargeDups	20	573	28.65	35.1885	5339682	0.0001
2	172	VP1201500374	CopyBacks	11	152	13.8182	38.6472	5339682	0
2	186	VP1201500423	Deletions	12	69	5.75	5.4333	3188103	0
2	186	VP1201500423	SmallDups	15	105	7	5.379	3188103	0
2	186	VP1201500423	LargeDups	13	79	6.0769	6.9222	3188103	0
2	186	VP1201500423	CopyBacks	13	42	3.2308	4.509	3188103	0

US EVD SURVIVORS:

Survivor:	Days post Onset:	Specimen Type:	Specimen Number:	Chimera Type:	#Unique Deletions:	#Reads Chimeric:	Avg. # Reads Per Chimera:	Standard Deviation Reads Per Chimera	Total # Mapped Reads:	Proportion Mapped Chimeric Reads:
A	28	semen	201403120	Deletions	1	4	4	0	145785	0
A	28	semen	201403120	SmallDups	2	33	16.5	1.5	145785	0.0002
A	28	semen	201403120	LargeDups	8	57	7.125	3.8871	145785	0.0004
A	58	semen	201403184	LargeDups	3	266	88.6667	62.4304	192147	0.0014
C	5	blood	201403131	Deletions	42	84	2	1.291	606711	0.0001
C	5	blood	201403131	SmallDups	45	123	2.7333	2.2549	606711	0.0002
C	5	blood	201403131	LargeDups	57	142	2.4912	1.6975	606711	0.0002
C	5	blood	201403131	CopyBacks	13	13	1	0	606711	0
C	7	blood	201403142	Deletions	71	102	1.4366	1.1596	583227	0.0002
C	7	blood	201403142	SmallDups	82	107	1.3049	0.7104	583227	0.0002
C	7	blood	201403142	LargeDups	134	175	1.306	0.7353	583227	0.0003
C	7	blood	201403142	CopyBacks	34	36	1.0588	0.3379	583227	0.0001
C	9	blood	201403147	Deletions	31	57	1.8387	0.8461	524944	0.0001
C	9	blood	201403147	SmallDups	68	129	1.8971	2.3272	524944	0.0002
C	9	blood	201403147	LargeDups	101	168	1.6634	0.9676	524944	0.0003
C	9	blood	201403147	CopyBacks	17	18	1.0588	0.2353	524944	0
C	12	blood	201403162	Deletions	13	23	1.7692	1.2499	91224	0.0003
C	12	blood	201403162	SmallDups	7	7	1	0	91224	0.0001
C	12	blood	201403162	LargeDups	33	39	1.1818	0.3857	91224	0.0004
C	12	blood	201403162	CopyBacks	1	1	1	0	91224	0
C	27	urine	201403234	Deletions	11	40	3.6364	3.7725	1201902	0
C	27	urine	201403234	SmallDups	17	452	26.5882	61.5296	1201902	0.0004
C	27	urine	201403234	LargeDups	38	590	15.5263	42.5989	1201902	0.0005
C	27	urine	201403234	CopyBacks	1	1	1	0	1201902	0
C	33	urine	201403258	SmallDups	1	7	7	0	877878	0
C	45	semen	201403360	Deletions	2	2	1	0	837349	0
C	45	semen	201403360	SmallDups	4	4254	1063.5	1832.2228	837349	0.0051
C	45	semen	201403360	LargeDups	4	4832	1208	2064.102	837349	0.0058
C	72	semen	201403439	Deletions	1	2	2	0	15350	0.0001
C	72	semen	201403439	SmallDups	1	4	4	0	15350	0.0003
C	72	semen	201403439	LargeDups	1	3	3	0	15350	0.0002
C	101	eye	201403522	Deletions	108	151	1.3981	0.8044	1264908	0.0001
C	101	eye	201403522	SmallDups	92	208	2.2609	3.4385	1264908	0.0002
C	101	eye	201403522	LargeDups	158	233	1.4747	1.8235	1264908	0.0002
C	101	eye	201403522	CopyBacks	2	3	1.5	0.5	1264908	0
C	117	semen	201403557	SmallDups	1	5	5	0	7479	0.0007
E	1	blood	201403368	Deletions	1	1	1	0	197384	0
E	1	blood	201403368	LargeDups	2	54	27	6	197384	0.0003
E	1	blood	201403368	CopyBacks	1	1	1	0	197384	0
E	2	plasma	201403391	Deletions	23	75	3.2609	2.6738	631517	0.0001
E	2	plasma	201403391	SmallDups	21	101	4.8095	4.0897	631517	0.0002
E	2	plasma	201403391	LargeDups	43	149	3.4651	2.5183	631517	0.0002
E	2	plasma	201403391	CopyBacks	8	8	1	0	631517	0
E	5	plasma	201403394	Deletions	5	136	27.2	18.7553	297798	0.0005
E	5	plasma	201403394	SmallDups	9	177	19.6667	19.1079	297798	0.0006
E	5	plasma	201403394	LargeDups	12	128	10.6667	10.3789	297798	0.0004
E	5	plasma	201403394	CopyBacks	1	1	1	0	297798	0
E	50	semen	201403509	Deletions	3	9	3	1.4142	58765	0.0002
E	50	semen	201403509	SmallDups	7	11	1.5714	0.9035	58765	0.0002
E	50	semen	201403509	LargeDups	8	12	1.5	1	58765	0.0002
E	50	semen	201403509	CopyBacks	1	2	2	0	58765	0

CELL CULTURE *in vitro* INFECTION

Time Point:	MOI:	Chimera Type:	#Unique Deletions:	#Reads Chimeric:	Avg. # Reads Per Chimera:	Standard Deviation Reads Per Chimera:	Total # Mapped Reads:	Proportion Mapped Chimeric Reads:
1 hr	2.0	Deletions	23	298	12.9565	48.7981	159328	0.0019
1 hr	2.0	SmallDups	22	32	1.4545	0.6556	159328	0.0002
1 hr	2.0	LargeDups	16	20	1.25	0.75	159328	0.0001
1 hr	2.0	CopyBacks	2	3	1.5	0.5	159328	0
18 hr	2.0	Deletions	150	2865	19.1	149.923	438530	0.0065
18 hr	2.0	SmallDups	76	87	1.1447	0.622	438530	0.0002
18 hr	2.0	LargeDups	79	95	1.2025	0.5126	438530	0.0002
18 hr	2.0	CopyBacks	58	58	1	0	438530	0.0001
1 hr	0.2	Deletions	4	13	3.25	3.3448	14540	0.0009
1 hr	0.2	SmallDups	3	6	2	1.4142	14540	0.0004
1 hr	0.2	LargeDups	1	1	1	0	14540	0.0001
48 hr	0.2	Deletions	101	581	5.7525	30.2145	401382	0.0014
48 hr	0.2	SmallDups	82	96	1.1707	0.5589	401382	0.0002
48 hr	0.2	LargeDups	99	105	1.0606	0.2386	401382	0.0003
48 hr	0.2	CopyBacks	143	149	1.042	0.2005	401382	0.0004

SUPPLEMENTARY EXPERIMENTAL METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding authors, Ute Ströher (ute.stroeher@gmail.com) and Gustavo Palacios (gustavo.f.palacios.ctr@mail.mil).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

Through the joint Sierra Leone Ebola Virus Persistence study (SLEVPS) with the Ministry of Health and Sanitation (MoHS) in Sierra Leone, WHO, China-CDC, and CDC, we had access to semen specimens collected from EVD survivors (Deen et al., 2015). Through this study we did not have access to direct patient data, such as patient age. Male study participates were stratified and selected for sequencing based on their NP Ct value and number/time span of serial semen specimens. As the SLEVPS only focused on specimen collection from EVD survivors, we did not have access to acute specimens from these participants. The SLEVPS was reviewed and approved by the Sierra Leone Institutional Review Board and the World Health Organization Ethical Review Committee. Following clinical diagnostic testing in the US, we did have access to paired acute blood and persistent semen specimens collected from US EVD patients. Acute and persistent specimens from US EVD survivors were collected by their treating physicians and transported to the CDC for detection of viral RNA (Kraft et al., 2015; Lyon et al., 2014; McElroy et al., 2015; Varkey et al., 2015). This sequencing project was determined by the CDC institutional human subject advisor to be a non-research public health response activity, and institutional review board review was not required.

METHOD DETAILS

Whole Genome Sequencing and Bioinformatics

RNA was extracted from blood and semen specimens using MagMAX Pathogen RNA/DNA isolation kit (Invitrogen) and BeadRetriever (Invitrogen) and treated with recombinant DNase I RNase-free (Roche). Ribosomal and carrier RNA were removed as previously described (Matranga et al., 2014). Non-depleted and rRNA/carrier RNA depleted specimens were prepared for sequencing using a modified version of the Illumina TruSeq RNA Access Library Prep kit as described previously with some minor variations (Blackley et al., 2016; Levin et al., 2010; Mate et al., 2015; Parkhomchuk et al., 2009; Sultan et al., 2012; Wang et al., 2011). RNA was fragmented for one minute prior to cDNA synthesis and custom dual indexes were used to avoid any sequencer bleed-through (Kircher et al., 2012). All specimens were enriched separately to avoid any bias of enriching one or two libraries over others in a pool. Specimens were sequenced using a Illumina MiSeq (version 3, 2x151 cycles), an Illumina Nextseq500 (midoutput kit, 2x151 cycles), and an Illumina HiSeq 2500 (rapid run v2, 2x151 cycles).

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). The priming sites of the random hexamer and Illumina TruSeq adaptors 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_qual_right 20 -min_len 50) (Schmeider, 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'). Genomes acquired from clinical specimens were deposited into Genbank: KY401638-KY401675, KY805810-2.

Analysis of Viral Evolutionary Rates

Viral evolutionary rate estimates were conducted using both linear regression modeling and time-structured phylogenies. For SAVS from SLE survivors, 1,058 EBOV genomes from Sierra Leone were analyzed using Path-O-Gen (now called TempEst (Rambaut, 2016)) and a maximum likelihood tree (GTR+G) rooted on the earliest available Sierra Leone sequence. For SAVS from US EVD survivors, 1498 genomes, representing a majority of sequences from Sierra Leone, Guinea and Liberia, were analyzed using Path-O-Gen (now called TempEst (Rambaut, 2016)) and a maximum likelihood tree (GTR+G) rooted on the earliest available Guinea sequence. Evolutionary rates and residual density plots were analyzed using R and custom python scripts from (Park et al., 2015). Evolutionary rate estimates for SAVS were also obtained using BEAST/v1.8.2,1.8.3, 1.8.4 (Drummond et al., 2012). A random selection of viral sequences, representing 25% of available sequences from SLE, or SLE/LBR/GIN, were used for the Bayesian analysis by partitioning into concatenated coding and noncoding sites. Rate estimates were modeled using unlinked HKY nucleotide evolutionary models with 4-independent Γ distributions, Bayesian skygrid demographic model (with variable population model estimated between January 1, 2014 and January 1, 2016, ie – "Time at last point:2"; or constant population, ie – "Time at last point:0"), and fixed local clock (Yoder and Yang, 2000) or uncorrelated lognormal local clock (Drummond et al., 2006) set with an initial prior of $1.1*10^{-3}$ subs/site/year. Model comparisons were conducted using: 1) relaxed uncorrelated lognormal clock with no constraints on the tree prior, variable Skygrid population; 1) relaxed uncorrelated lognormal clock with no constraints on the tree prior, constant population: "$UCLN_{unconstrained}$"; 2) relaxed uncorrelated lognormal clock with individual survivor blood and/or semen sequences constrained to survivor-specific monophyletic blood/semen taxons, constant population: "$UCLN_{monophyletic}$", 3) Fixed local clock with individual survivor blood and/or semen sequences constrained to survivor-specific monophyletic blood/semen taxons, constant population: and 4) Fixed local clock with survivor blood and/or semen sequences

constrained to blood-specific and semen-specific taxons, constant population: "$FLC_{monophyletic}$". The MCMC analysis was conducted for 800 million generations, which represents a compilation of 8-independent replicates of 100 million generations (sampled every $10{,}000^{th}$ state). Convergence was obtained for the majority of replicates and burn-in was removed (usually 5-10% of total states) by examining the trace and effective sample size statistics (min ESS > 200 for all models) using tracer/v1.6. Strength of model fit was evaluated by performing path- and stepping stone-sampling with default values and best-of-fit was evaluated by calculating Bayes Factors. Survivor and acute rate estimates from Bayesian analysis conducted with the UCLN clock models were estimated using custom-modified samogitia.py scripts (Dudas, 2017).

Sequence Analysis

Additional sequence analysis was conducted using CLC Genomics/v9.0. Potential hyper-edited sites due to host-encoded adenosine deaminases acting on RNA (ADARs) can result in the rapid accumulation of clustered T(U)-to-C substitutions (on the positive strand) in the EBOV genome (Dudas et al., 2017). We identified clusters of substitutions consistent with ADAR-mediated editing (≥3 phylogenetically-linked T(U)-to-C substitutions within a 200 nt window), and these substitutions were masked for evolutionary rate analyses (i.e., C genotypes were converted to T at these positions) in Figure 2A-B, and Figure 3A-B. Histograms of U-to-C hyper-editing were generated using R. Median joining networks were constructed using sequence alignments from each EVD survivor with PopART/v1.7.2. Intrahost variants (iSNVs) were detected with FreeBayes v1.0.2 (Garrison, 2012). For iSNV detection, we only used reads with mapping quality ≥30 and positions with base quality ≥30. An iSNV was only considered if the alternate allele was represented by ≥5 reads and present at a frequency ≥3%. We estimated SNV and insertion frequencies for the longitudinal phasing analysis by first performing a read-pair merging of the assemblies in IRMA v0.6.5 (Shepard et al., 2016) and computing allele frequencies for each selected site using IRMA's *call.pl* script (–B option). A pairwise (Manhattan)

distance matrix was computed in R v3.3.1 for each position-allele combination with the vector of the observed frequencies ordered by specimen date. The matrix was used to generate a single linkage dendrogram, also in R. SNVs were divided into two clusters based on two near-symmetrical branches in the tree. The insertion frequency of C at upstream position 6924 and its complement were added to the dataset and a second dendrogram produced. The 6924 C-insert and 6924 non-C-insert frequencies were assigned an SNV cluster according to their nearest neighboring SNV in the second tree: 8371G and 8371A respectively. A final tree and distance matrix was produced for each variant position by ordering the frequency vectors by specimen date as well as variant cluster (one allele or insertion state was assigned to each cluster for each site). Frequency line graphs of positions, alleles, and specimen dates were created using Tableau v10.0 and the positions in the graph ordered and composited with the final R dendrogram in Supplementary Figure 3. Identification of chimeric reads were performed by mapping 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) using bwa and chimeric reads were further defined using custom scripts. US EVD survivors received multiple thereapeutic treatments (A: whole blood transfusion, convalescent whole blood transfusion, ZMAPP; C: TKM-Ebola, convalescent plasma; E: Convalescent plasma, brincidofovir)(Kraft et al., 2015; Lyon et al., 2014; McElroy et al., 2015; Varkey et al., 2015) and we confirmed that viral regions targeted by these compounds (GP, VP35, polymerase) did not mutate through comparison of serial consensus viral sequences.

Estimation of Selective Pressure

Evolutionary selective pressures were estimated using the renaissance counting method in beast/v1.8.2 (Lemey et al., 2012; O'Brien et al., 2009) with a subset of genomes representing 25% of available random sequences from Sierra Leone, Guinea, and Liberia that did not contain codon frame shifts.

123 Codon alignments for each gene were partitioning into coding and concatenated total noncoding sites.
124 Rate estimates were modeled using unlinked HKY nucleotide substitution models, Bayesian skygrid
125 demographic model, and uncorrelated lognormal relaxed clock set with an initial prior of $1.1*10^{-3}$
126 subs/site/year. The MCMC analysis was conducted for 400 million generations for each gene, which
127 represents a compilation of 4-independent replicates of 100 million generations (sampled every 1000^{th}
128 state). Due to time constraints, MCMC analysis for the VP40 and polymerase gene were stopped at
129 ~200 million or ~120 million iterations, which easily reached convergence. For all replicates,
130 convergence was obtained and burn-in was removed (usually 10% of total states) by examining the trace
131 and effective sample size statistics (>200 for all MCMC analyses) using tracer/v1.6. Only one MCMC
132 replicate for the CGP tail did not converge, and it was removed from additional analysis. Omega
133 estimates were calculated by using the conditioned and unconditioned N and S estimates and equation 1
134 ((total_N/total_S) / (unconditioned_N/unconditioned_S)) from Lemey *et al.* (Lemey et al., 2012) and
135 scripts from Park *et al*. (Park et al., 2015). To prevent rate overestimation by double-counting shared
136 amino acids, the glycoprotein was split at the transcriptional editing site (nucleotide 6923) into N-
137 terminal (nucleotides 6039-6923, "NGP"), C-terminal full length (nucleotides 6923-8068 - containing
138 the GP1 carboxy-terminus and GP2, "CGP") and secreted GP (nucleotides 6924-7157, "SGP_c"). For
139 secreted GP (nucleotides 6924-7157, "SGP_c") rate estimates, approximately 9.6% of unconditioned S
140 estimates and 0.2% of unconditioned N estimates were 0.0; thus to bypass undefined ω estimates these
141 values were converted to 1. For polymerase rate estimates, approximately 3% of N or S estimates were
142 undefined (NaN) and to bypass undefined ω estimates these states were removed from the analysis.
143
144 Selective pressure hypothesis testing was performed using the codeml model in paml/v4.5 with a subset
145 of 231 genomes, representing approximately 25% of available random sequences from Sierra Leone,
146 Guinea and Liberia that did not contain reading frame shifts. We constructed a Maximum Clade

147 Credibility tree using beast/v.1.8.2 by partitioning the alignments into concatenated coding and
148 noncoding sites and trees were modeled using unlinked HKY nucleotide substitution models, Bayesian
149 skygrid demographic model, and uncorrelated lognormal relaxed clock set with an initial prior of
150 $1.1*10^{-3}$ subs/site/year. The MCMC analysis was conducted for 50 million generations (sampled every
151 1000^{th} state), which easily reached convergence. The cladogram of the MCC tree was used as input for
152 paml codeml. Branch model testing was performed using model0 and model2 and branch-site testing
153 was performed using modelA and A_null with codon frequencies F3x4. For branch testing, kappa and
154 omega estimates from model0 were set as initial estimates for model2 (acute sequences vs. SAVS) and
155 model2 (acute sequences vs. SAVS_acute_rate vs. SAVS_slow_rate). Strength of statistical support for
156 models2 (alternative hypotheses) vs. model0 (null hypothesis) was measured using the 2Δlog-likelihood
157 method with degrees of freedom=1 and further corrected according to Bonferroni (p = 0.05/ 2 tests
158 conducted with same sequence alignment) (Anisimova and Yang, 2007; Yang, 2007). The ratio of N to
159 S was calculated by summing the total N (N*dN from PAML model2 output) and S (S*dS from PAML
160 model2 output) estimates for all acute and SAVS branches and dividing by the total N and S count. For
161 branch-site testing, semen-specific branches were set as foreground branches and modelA was
162 performed using NSsites=2 with kappa and omega estimates set at initial values from model0.
163 ModelA_null testing was performed with NSsites=2, kappa and omega estimates set at initial values
164 from model0, and omega fixed at 1. Significance values were calculated using the 2Δlog-likelihood
165 method and significance was established with p values below 0.05.

166

167 *Ebola virus* in vitro *Infection*

168 All work with EBOV was performed in a biosafety level 4 (BSL-4) facility. Huh7 cells were cultured in
169 Dulbecco's modified Eagle's medium high glucose (DMEM) (item number 11960-044, Invitrogen)
170 supplemented with 10% heat inactivated HyClone fetal bovine serum (Thermo Scientific), 1x non-

essential amino acids (Invitrogen), 1x penicillin-streptomycin (Invitrogen), and 1x Glutamax (Invitrogen) at 37°C with 5% CO_2. Prior to viral infection, cells were seeded into triplicate wells in a 12-well plate and media was replaced with FluorBrite Dulbecco's modified Eagle's medium (DMEM) (item number A1896701, Invitrogen) supplemented with 10% heat inactivated HyClone fetal bovine serum (Thermo Scientific), 1x non-essential amino acids (Invitrogen), 1x penicillin-streptomycin (Invitrogen), and 1x Glutamax (Invitrogen). Immediately before infection cells were counted using the Moxi Z cell counter with M cassettes (Orflo, Technologies). Cells were infected at MOI of 2 or 0.2 with rEBOV-L2014/ZsG (Albarino et al., 2016) in 200uL of FluorBrite media with supplements for 1 hour at 37°C with 5% CO_2. After absorption, inoculum was removed and cells were washed twice with 1mL of PBS. Media was replaced with FluorBrite media for infection duration. At specified time points supernatant and monolayers were inactivated with TriPure isolation reagent (Roche). Prior to inactivation supernatants were spun at 200xg for 10 minutes to remove cellular debris. RNA was extracted from Tripure using the Direct-zol-96 MagBead RNA isolation kit (Zymo Research). Active viral infection of cells was confirmed by visualization of ZsGreen fluorescence at 1, 18, and 48 hours post infection.

Ebola virus RNA Strandedness Analysis

The TruSeq RNA Access Library Prep kit results in stranded data (i.e., read 1 is complementary to the original RNA molecule). Using custom scripts we quantified the proportion of positive- and negative-sense RNA molecules present in each specimen. Independently for each strand and each specimen, we also calculated relative depth of coverage for every EBOV ORF as

$$\widehat{D_j} = \frac{\frac{1}{n_j}\sum_{i=1}^{n_j} D_{ji}}{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{n_j}\sum_{i=1}^{n_j} D_{ji}}$$

N is the total number of ORFs, n_j is the length in nucleotides of the j[th] ORF and D_{ji} is the read 1 depth of coverage at the i[th] nucleotide of the j[th] ORF. Only read 1 was used to avoid double counting at positions where reads 1 and 2 overlap. Regions of the genome included in multiple ORFs were excluded. In Figure 4, strandedness data from next generation sequencing is presented either for 1) specific genes (Figures 4D, Supplementary Figures 5F and 5G), 2) whole genome sequences (Figures 4A-C, and Supplementary Figures 5A-C), or 3) for the NP-specific open reading frame (Supplementary Figures 6A-D). Cutoff criteria was ≥50x average paired end coverage across the genome for Figures 4A-C and Supplementary Figures 5A-C, ≥25x average read 1 coverage across the coding portions of the genome (Figure 4D and Supplementary Figure 5F), and ≥25x average read 1 coverage across NP (Supplementary Figure 6A-D).

The EBOV NP strand-specific qRT-PCR assay was performed by using separate first and second strand reactions. The first-strand reaction was conducted with 2.5uL of input RNA, 1uL 10mM dNTPs (Invitrogen), 1uL of 2uM gene-specific tagged stranded primer and 5uL of nuclease-free water (Ambion). This mixture was heated to 65°C for 5 minutes and placed on ice for 2 minutes. The reverse transcription reaction followed with 4uL of 5x first-strand reaction buffer (Invitrogen), 1uL SUPERase-In (Invitrogen), 1uL superscript III reverse transcriptase (Invitrogen), 1uL 0.1M DTT (Invitrogen) and 3.5 uL of nuclease-free water (Ambion). The reaction was heated at 55°C for 15 minutes and cooled on ice for 2 minutes. First strand reactions were cleaned with the QiaQuick PCR cleanup kit (Qiagen) and ssDNA was eluted with 30uL of nuclease-free water (Ambion). The second strand reaction proceeded with 5uL of input cDNA, 2.5uL of AmpliTaq 10x buffer I, 0.5uL of 10mM dNTP's (Invitrogen), 2.25uL of 10uM tag-specific primer, 2.25uL of 10uM gene-specific primer, 0.625uL of 10uM NP probe, 0.125uL of AmpliTaq DNA polymerase (Invitrogen), and 11.75uL of nuclease-free water (Ambion). Thermocycler conditions consisted of 50°C for 15 minutes, 95°C for 2 minutes, 95°C for 15s and 55°C for 45s (44 cycles). To convert Ct values into strand copy numbers, we established a Ct versus molarity

concentration curves for both positive- and negative-sense synthetic RNA's. Goodness-of-fit values for these curves (r^2) were all greater than 0.988. Using the same first-strand cDNA products, we also established Ct versus copy number using the Bio-Rad QX200 digital droplet PCR and r^2 values for these curves were all greater than 0.979. Reaction conditions for ddPCR consisted of 10uL of 2x ddPCR Supermix for Probes (Bio-Rad), 1.8uL of 10uM tag-specific primer, 1.8uL of 10uM gene-specific primer, 0.5uL of 10 uM NP-specific probe, uL of cDNA, and 0.9uL of nuclease-free water (Ambion). Thermocycler conditions consisted of 95°C for 10 minutes, 94°C for 30s, 60°C for 1 minute (39 cycles), and 98°C for 10 minutes with a ramp speeds done at 2°C/sec. Final Ct to copy number conversions for *in vitro* infections and EVD survivors clinical specimens were calculated using the Ct versus molarity concentration curves corrected for copy numbers as estimated using ddPCR.

QUANTIFICATION AND STATISTICAL ANALYSIS

For evolutionary rate estimates using Bayesian analysis we present the mean and 95% highest posterior density estimates calculated from all total combined states (after removal of burn-in, in most cases 10%) using scripts from Park et al. and custom-modified samogitia.py scripts (Dudas, 2017; Park et al., 2015). Evolutionary rates estimates from RTT's are presented as the line of best fit with 95% confidence intervals shaded in grey. Residual comparisons from linear regressions display the 2-fold standard deviations of the acute residual density in grey. Strength of statistical support for paml estimation of selective pressure was measured using the likelihood ratio test with degrees of freedom=1 comparing model0 (null hypothesis) with model2 (alternative hypotheses). Significance values for modelA and modelA_null branch-site testing with PAML were calculated using the 2Δlog-likelihood method and significance was established with p values below 0.05. A one-way analysis of variance (ANOVA) for the association of proportion of positive-sense reads or NP Ct values vs. virus isolation result was assessed using the lm() and anova() functions from R v3.3.1.

DATA AND SOFTWARE AVAILABILITY

Most software utilized is freely available, and when possible we include the version number and reference for the software used. Custom scripts have been submitted to github (https://github.com/jtladner/Scripts and https://github.com/evk3/EBOV_semen_sequencing). Genomes acquired from clinical specimens were deposited into Genbank: KY401638-KY401675 and KY805810-2.

REFERENCES:

Albarino, C.G., Guerrero, L.W., Chakrabarti, A.K., Kainulainen, M.H., Whitmer, S.L., Welch, S.R., and Nichol, S.T. (2016). Virus fitness differences observed between two naturally occurring isolates of Ebola virus Makona variant using a reverse genetics approach. Virology *496*, 237-243.

Anisimova, M., and Yang, Z. (2007). Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol *24*, 1219-1228.

Blackley, D.J., Wiley, M.R., Ladner, J.T., Fallah, M., Lo, T., Gilbert, M.L., Gregory, C., D'Ambrozio, J., Coulter, S., Mate, S.*, et al.* (2016). Reduced evolutionary rate in reemerged Ebola virus transmission chains. Sci Adv *2*, e1600378.

Deen, G.F., Knust, B., Broutet, N., Sesay, F.R., Formenty, P., Ross, C., Thorson, A.E., Massaquoi, T.A., Marrinan, J.E., Ervin, E.*, et al.* (2015). Ebola RNA Persistence in Semen of Ebola Virus Disease Survivors - Preliminary Report. The New England journal of medicine.

Drummond, A.J., Ho, S.Y., Phillips, M.J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS Biol *4*, e88.

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.

Dudas, G. (2017). BALTIC - adaptable lightweight tree import code for molecular phylogeny manipulation, analysis and visualisation (github.com).

Dudas, G., Carvalho, L.M., Bedford, T., Tatem, A.J., Baele, G., Faria, N.R., Park, D.J., Ladner, J.T., Arias, A., Asogun, D.*, et al.* (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature *544*, 309-315.

Garrison, E.a.M., G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv *1207.3907 [q-bio.GN]*

Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S., 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.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res *40*, e3.

Kraft, C.S., Hewlett, A.L., Koepsell, S., Winkler, A.M., Kratochvil, C.J., Larson, L., Varkey, J.B., Mehta, A.K., Lyon, G.M., 3rd, Friedman-Moraco, R.J.*, et al.* (2015). The Use of TKM-100802 and Convalescent Plasma in 2 Patients With Ebola Virus Disease in the United States. Clin Infect Dis *61*, 496-502.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lemey, P., Minin, V.N., Bielejec, F., Kosakovsky Pond, S.L., and Suchard, M.A. (2012). A counting renaissance: combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. Bioinformatics *28*, 3248-3256.
Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods *7*, 709-715.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Lyon, G.M., Mehta, A.K., Varkey, J.B., Brantly, K., Plyler, L., McElroy, A.K., Kraft, C.S., Towner, J.S., Spiropoulou, C., Stroher, U.*, et al.* (2014). Clinical care of two patients with Ebola virus disease in the United States. The New England journal of medicine *371*, 2402-2409.
Martin, M. (2011). Cutadapt removes adaptor sequences from high-throughput sequences reads. EMBnetjournal *17*, 10-12.
Mate, S.E., Kugelman, J.R., Nyenswah, T.G., Ladner, J.T., Wiley, M.R., Cordier-Lassalle, T., Christie, A., Schroth, G.P., Gross, S.M., Davies-Wayne, G.J.*, et al.* (2015). Molecular Evidence of Sexual Transmission of Ebola Virus. The New England journal of medicine *373*, 2448-2454.
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.
McElroy, A.K., Akondy, R.S., Davis, C.W., Ellebedy, A.H., Mehta, A.K., Kraft, C.S., Lyon, G.M., Ribner, B.S., Varkey, J., Sidney, J.*, et al.* (2015). Human Ebola virus infection results in substantial immune activation. Proc Natl Acad Sci U S A *112*, 4719-4724.
O'Brien, J.D., Minin, V.N., and Suchard, M.A. (2009). Learning to count: robust estimates for labeled distances between molecular sequences. Mol Biol Evol *26*, 801-814.
Park, D.J., Dudas, G., Wohl, S., Goba, A., Whitmer, S.L., Andersen, K.G., Sealfon, R.S., Ladner, J.T., Kugelman, J.R., Matranga, C.B.*, et al.* (2015). Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. Cell *161*, 1516-1526.
Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res *37*, e123.
Rambaut, A., Lam, T.T., Carvalho, L.M., Pybus, O.G. (2016). Exploring the temporal structure of heterchronus sequences using TempEst (formerly Path-O-Gen). Virus Evolution, vew007.
Schmeider, R.a.E., R (2011). Quality Control and preprocessing of metagenomics datasets. Bioinformatics *27*, 863-864.
Shepard, S.S., Meno, S., Bahl, J., Wilson, M.M., Barnes, J., and Neuhaus, E. (2016). Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. BMC Genomics *17*, 708.
Sultan, M., Dokel, S., Amstislavskiy, V., Wuttig, D., Sultmann, H., Lehrach, H., and Yaspo, M.L. (2012). A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. Biochem Biophys Res Commun *422*, 643-646.
Varkey, J.B., Shantha, J.G., Crozier, I., Kraft, C.S., Lyon, G.M., Mehta, A.K., Kumar, G., Smith, J.R., Kainulainen, M.H., Whitmer, S.*, et al.* (2015). Persistence of Ebola Virus in Ocular Fluid during Convalescence. The New England journal of medicine.
Wang, L., Si, Y., Dedow, L.K., Shao, Y., Liu, P., and Brutnell, T.P. (2011). A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. PLoS One *6*, e26426.
Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*, 1586-1591.
Yoder, A.D., and Yang, Z. (2000). Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol *17*, 1081-1090.

329 Materials Table:

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Dnase I, Rnase-free	Roche	4716728001
Hybridase, Rnase H	Epicentre	H39100
Qiagen Rnase-free Dnase I	Qiagen	79254
Superase-in	Ambion	AM1694
Agencourt RNAclean XP SPRI clean up beads	Beckman Coulter Genomics	41105518
Dulbecco's modified Eagle's medium high glucose (DMEM)	Invitrogen	11960-044
HyClone fetal bovine serum	Thermo Scientific	SH30070.03
100x Non-essential amino acids	Invitrogen	11140050
10,000 U/mL Penicillin-streptomycin	Invitrogen	15140122
GlutaMAX	Invitrogen	35050061
FluoroBrite FluorBrite Dulbecco's modified Eagle's medium (DMEM)	Invitrogen	A1886701
TriPure Isolation reagent	Roche	11667165001
Critical Commercial Assays		
MagMax Pathogen RNA/DNA Isolation Kit	Invitrogen	4462359
Direct-zop-96 MagBead RNA Isolation Kit	Zymo Research	
TruSeq RNA Access Library Prep kit	Illumina	RS-301-2001
MiSeq Reagent Kit vs (300 cycle)	Illumina	MS-102-2002
HiSeq Rapid PE Cluster Kit v2	Illumina	PE-402-4002
TruSeq Rapid Duo cBot v1 Sample loading kit	Illumina	CT-402-4001
Moxi Z cell counter, M cassettes	Orflo Technologies	MXC001
Deposited Data		
EBOV genomcs from clinical specimens	This study	KY401638-KY401675, KY805810-2

Experimental Models: Cell Lines		
Huh7 cells	Albarino, C.G., Guerrero, L.W., Chakrabarti, A.K., Kainulainen, M.H., Whitmer, S.L., Welch, S.R., and Nichol, S.T. (2016). Virus fitness differences observed between two naturally occurring isolates of Ebola virus Makona variant using a reverse genetics approach. Virology 496, 237-243.	
Experimental Models: Organisms/Strains		
rEBOV-L2014/ZsG	Albarino, C.G., Guerrero, L.W., Chakrabarti, A.K., Kainulainen, M.H., Whitmer, S.L., Welch, S.R., and Nichol, S.T. (2016). Virus fitness differences observed between two naturally occurring isolates of Ebola virus Makona variant using a reverse genetics approach. Virology 496, 237-243.	
Sequence-Based Reagents		
rRNA-specific DNA probes	Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ: Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods. 2013, 10: 623-629. 10.1038/nmeth.2483.	
Oligo dT, 40 nucleotides	idtdna.com	
Ebola virus-specific custom hybridization probes	Mate, S.E., Kugelman, J.R., Nyenswah, T.G., Ladner, J.T., Wiley, M.R., Cordier-Lassalle, T., Christie, A., Schroth, G.P., Gross, S.M., Davies-Wayne, G.J., et al. (2015). Molecular Evidence of Sexual Transmission of Ebola Virus. The New England journal of medicine 373, 2448-2454.	
Positive-sense NP synthetic RNA	This study.	Available upon request.
Negative-sense NP synthetic RNA	This study.	Available upon request.
NP forward_tagged strand-spcific primer	This study.	Available upon request.
NP reverse_tagged strand-specific primer	This study.	Available upon request.
Forward tagged primer	This study.	Available upon request.
Reverse tagged primer	This study.	Available upon request.
NP-specific probe	This study.	Available upon request.
Software and Algorithms		
Cutadapt v1.21	Martin, M. (2011). Cutadapt removes adaptor sequences from high-throughput sequences reads. EMBnetjournal 17, 10-12.	

Prinseq-lite v0.20.4	Schmeider, R.a.E., R (2011). Quality Control and preprocessing of metagenomics datasets. Bioinformatics 27, 863-864.	
Bowtie2	Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359.	
Picard	broadinstitute.github.io/picard	
Samtools v0.1.18	Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.	
Path-O-Gen (now called TempEst)	Rambaut, A., Lam, T.T., Carvalho, L.M., Pybus, O.G. (2016). Exploring the temporal structure of heterchronus sequences using TempEst (formerly Path-O-Gen). Virus Evolution, vew007.	
R v3.3.1	R Core Team RFfSC: R: A Language and Environment for Statistical Computing. In. Vienna, Austria; 2012.	
Custom Python scripts	Park, D.J., Dudas, G., Wohl, S., Goba, A., Whitmer, S.L., Andersen, K.G., Sealfon, R.S., Ladner, J.T., Kugelman, J.R., Matranga, C.B., et al. (2015). Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. Cell 161, 1516-1526.	
Beast v1.8.2	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.	
Tracer v1.6	http://tree.bio.ed.ac.uk/software/tracer/	
CLC Genomics v9.0		
PopArt v1.7.2	http://popart.otago.ac.nz/index.shtml	
FreeBayes v1.0.2	Garrison, E.a.M., G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv 1207.3907 [q-bio.GN]	
IRMA v0.6.5	Shepard, S.S., Meno, S., Bahl, J., Wilson, M.M., Barnes, J., and Neuhaus, E. (2016). Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. BMC Genomics 17, 708.	
Tableau v10.0		
paml v4.5	http://abacus.gene.ucl.ac.uk/software/paml.html	
Custom strandedness, ADAR, and samogitia.py scripts	https://github.com/jtladner/ and https://github.com/evk3/EBOV_semen_sequencing	
Samogitia.py scripts	https://github.com/blab/baltic, by Gytis Dudas	

330

Sierra Leone Ebola Virus Persistence Study Group Affiliations:

- **Sierra Leone Ministry of Health and Sanitation:** Gibrilla Fadlu Deen (principle investigator), James Bangura, Amara Jambai, Faustine James, Alie H. Wurie, Francis Yamba
- **Sierra Leone Ministry of Defense:** Foday Sahr, Foday R. Sesay, Thomas A. Massaquoi
- **Sierra Leone Ministry of Social Welfare, Gender, Children's Affair:** Tina Davies
- **World Health Organization (WHO):** Nathalie Broutet (principle investigator), Pierre Formenty, Anna E. Thorsen, Archchun Ariyarajah, Marylin Carino, Antoine Coursier, Kara N. Durski, Ndema Habib, Philippe Gaillard, Sihem Landoulski, Margaret O. Lamunu, Jacyln E. Marrinan, Suzanna L.R. McDonald, Dhamari Naidoo
- **United States Centers for Disease Control and Prevention (US-CDC):** Barbara Knust (principle investigator), Neetu Abad, Kyle T. Bernstein, Elizabeth Ervin, John D. Klena, Tasneem Malik, Oliver Morgan, Stuart T. Nichol, Christine Ross, Ute Ströher
- **Chinese Center for Disease Control and Prevention (China-CDC):** Wnbo Xu (principle investigator), Hongtu Liu, William Jun Liu, Yong Xhang, Guizhen Wu, Mifang Liang
- **Joint United Nations Programme on HIV/AIDS (UNAIDS):** Patricia Ongpin