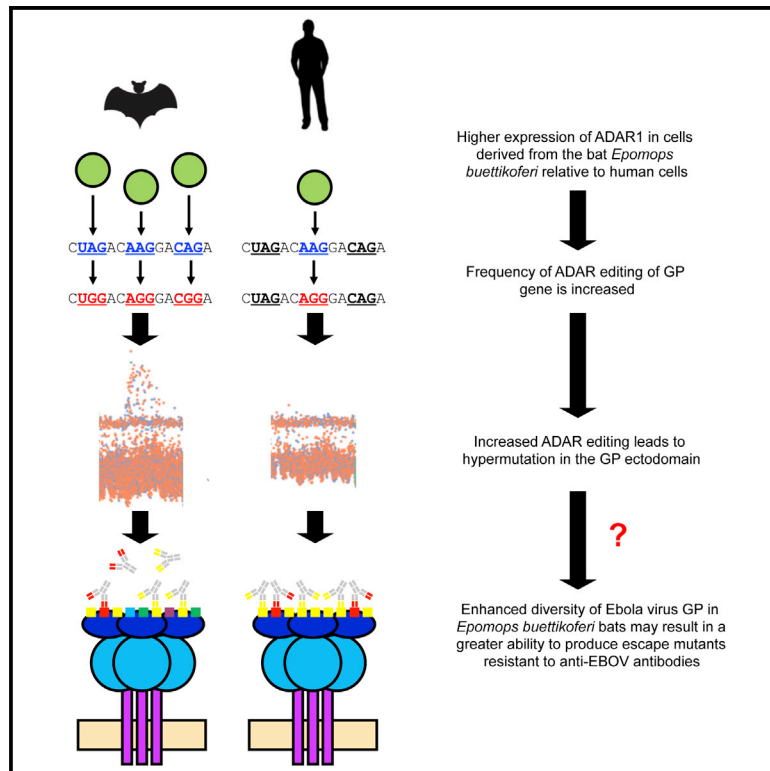# Species-Specific Evolution of Ebola Virus during Replication in Human and Bat Cells

Graphical Abstract

Authors

Zachary J. Whitfield, Abhishek N. Prasad, Adam J. Ronk, Ivan V. Kuzmin, Philipp A. Ilinykh, Raul Andino, Alexander Bukreyev

Correspondence

raul.andino@ucsf.edu (R.A.), alexander.bukreyev@utmb.edu (A.B.)

In Brief

Whitfield et al. use circular RNA sequencing to demonstrate that host species identity has effects on Ebola virus evolution and that the host RNA editing enzyme ADAR appears to edit the Ebola virus glycoprotein gene in bat cells.

Highlights

- Virus passaged in human cells exhibits adaptive mutations

- Bat cell-derived populations are influenced by host factors

- ADAR appears to edit the GP gene in bat cells

- Additional residues in EBOV ORFs influence replication

Article

Species-Specific Evolution of Ebola Virus during Replication in Human and Bat Cells

Zachary J. Whitfield,[1,5] Abhishek N. Prasad,[2,4,5] Adam J. Ronk,[2,4,5] Ivan V. Kuzmin,[2,4,5] Philipp A. Ilinykh,[2,4] Raul Andino,[1,*] and Alexander Bukreyev[2,3,4,6,*]

[1]Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA
[2]Department of Pathology, University of Texas Medical Branch, Galveston, TX, USA
[3]Department Microbiology & Immunology, University of Texas Medical Branch, Galveston, TX, USA
[4]Galveston National Laboratory, University of Texas Medical Branch, Galveston, TX, USA
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: raul.andino@ucsf.edu (R.A.), alexander.bukreyev@utmb.edu (A.B.)
https://doi.org/10.1016/j.celrep.2020.108028

SUMMARY

Ebola virus (EBOV) causes a severe, often fatal disease in humans and nonhuman primates. Within the past decade, EBOV has caused two large and difficult-to-control outbreaks, one of which recently ended in the Democratic Republic of the Congo. Bats are the likely reservoir of EBOV, but little is known of their relationship with the virus. We perform serial passages of EBOV in human and bat cells and use circular sequencing to compare the short-term evolution of the virus. Virus populations passaged in bat cells have sequence markers indicative of host RNA editing enzyme activity, including evidence for ADAR editing of the EBOV glycoprotein. Multiple regions in the EBOV genome appear to have undergone adaptive evolution when passaged in bat and human cells. Individual mutated viruses are rescued and characterized. Our results provide insight into the host species-specific evolution of EBOV and highlight the adaptive flexibility of the virus.

INTRODUCTION

Ebola virus (EBOV) is the prototypic virus of the genus *Ebolavirus*, a group of viruses associated with severe and frequently fatal disease in humans and nonhuman primates (NHP). The genus is a member of the family Filoviridae (order Mononegavirales), which includes the genera *Marburgvirus*, *Cuevavirus*, *Striavirus*, and *Thamnovirus* (Kuhn et al., 2019; Negredo et al., 2011; Yang et al., 2019). EBOV has caused the two largest filovirus outbreaks in recorded history: the West African epidemic of 2013–2016 (Agua-Agum et al., 2016) and the recent outbreak in the Democratic Republic of the Congo (DRC), which began in August 2018 and was contained only with substantial effort (Médecins sans Frontières, 2020; World Health Organization, 2020). The Egyptian fruit bat (*Rousettus aegyptiacus*), a megabat native to much of sub-Saharan Africa, has been definitively identified as a reservoir for viruses of the genus *Marburgvirus* (Towner et al., 2009), and strong evidence indicates that bats serve as the primary reservoir for EBOV as well (Goldstein et al., 2018; Leroy et al., 2005; Marí Saéz et al., 2015; Olival and Hayman, 2014; Taylor et al., 2011). Of particular note, EBOV RNA has been detected in bats of four species, *Hypsignathus monstrosus*, *Epomops franqueti*, *Myonycteris torquata*, and *Miniopterus inflatus* (Leroy et al., 2005; EcoHealth Alliance, 2019).

Like most RNA viruses, filoviruses encode a non-proofreading RNA-dependent RNA polymerase (RdRP). Consequently, genomic replication is far more error prone than in other organisms, resulting in higher mutation rates (Holmes, 2009). RNA virus genomes therefore face strong selective pressure to exhibit a significant degree of mutational robustness (Lauring et al., 2013). Another consequence is their remarkable ability to adapt to new replicative environments (Andino and Domingo, 2015). RNA virus replication produces complex population structures in which the replication of a single "master genome" (the consensus sequence) gives rise to a large, complex, and interconnected "mutant swarm" of variant genomes of varying degrees of fitness relative to the master genome. The impact of intra-host genetic diversity on virulence and fitness within the host is well documented for numerous RNA viruses, including hepatitis C virus (Farci et al., 2000), several enteroviruses (Meng and Kwang, 2014; Pfeiffer and Kirkegaard, 2005; Vignuzzi et al., 2005), chikungunya virus (Coffey et al., 2011), and West Nile virus (Grubaugh et al., 2015, 2016), in which reduced diversity of virus populations results in lower fitness and an attenuated infection phenotype. Mutation rates of RNA viruses are difficult to determine, but are estimated at the order of 10^{-6}–10^{-4} substitutions/nucleotide/cycle of replication (Holmes, 2009; Peck and Lauring, 2018). Although the mutation rate of EBOV is not firmly established, the evolutionary rate of the virus in humans (the rate at which genetic variants arise and proliferate throughout a virus population) is estimated to be $\sim$4.7 $\times$ 10^{-4} substitutions/site/year when averaged across all outbreaks from 1976 to 2018 (Mbala-Kingebeni et al., 2019). However, this figure is not directly comparable with mutation rate, as multiple factors, including population size and demographic
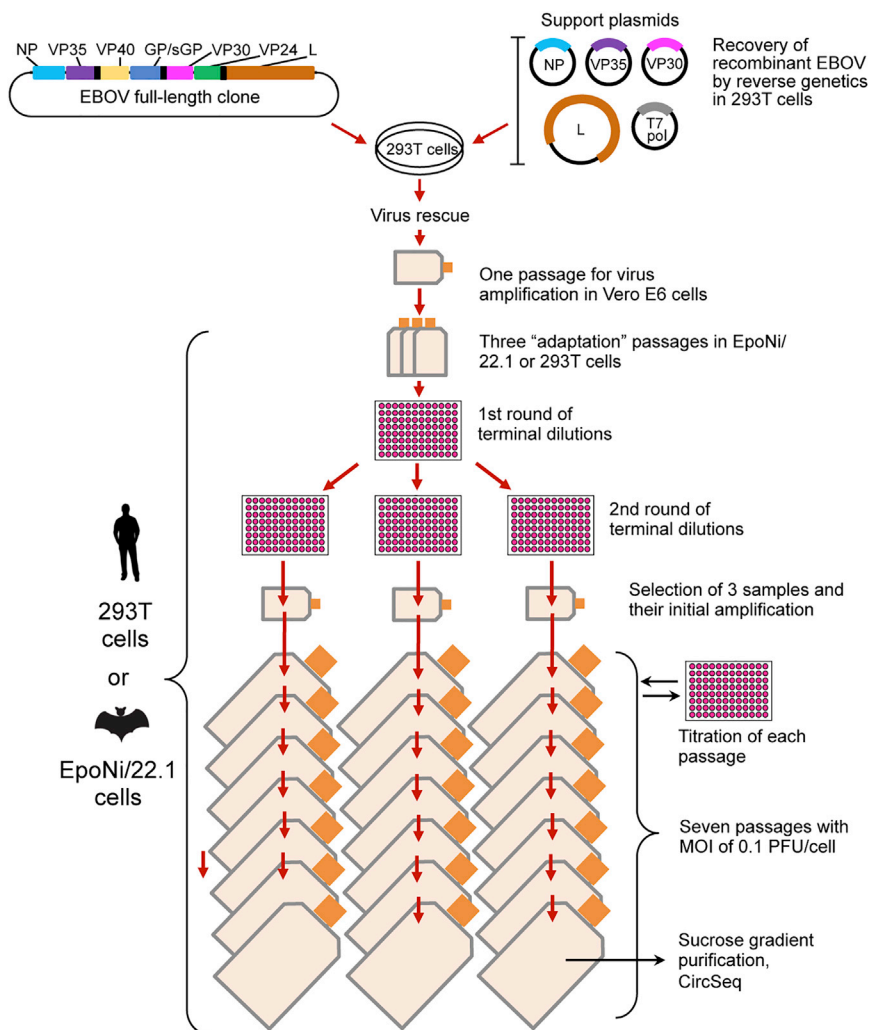
Figure 1. Schematic of Passaging Experimental Design

However, mice do not recapitulate human or NHP disease, and the size of the rhesus macaque groups used was insufficient to detect a possible shift in pathogenicity. Furthermore, no significant attempt was made to determine any effect of the mutants on transmission, a significant contributor to the fitness of a virus during an outbreak. In the present study, we sought to characterize EBOV adaptation to cells of bat and human origin. In order to assess changes in mutation rates and the structure of EBOV populations during serial passage through either human (293T) or bat (EpoNi/22.1, *Epomops buettikoferi*) renal cell lines, we used circular sequencing (CirSeq) (Acevedo et al., 2014). CirSeq is an Illumina platform-based ultra-deep-sequencing approach that uses specialized library preparation and computational protocols to eliminate the vast majority of sequencing errors, reducing the error rate of sequencing to as low as 10^{-12} per base. This permits variant calls at a far lower threshold. We identified differences in individual nucleotide mutation rates, as well as observed a number of host-specific mutations that appeared to have undergone positive selection. In addition, a particularly prominent cluster of mutations in the region spanning the glycan cap (GC) and mucin-like domain (MLD) of the glycoprotein (GP) of EBOV passaged in EpoNi/22.1 cells was identified. Finally, we selected several mutants from each cell line for further investigation using both infectious EBOV prepared via reverse genetics and the EBOV minigenome (MG) system. Along with characterization of replication kinetics in each cell line, co-infection experiments were performed to assess the fitness of the selected mutant viruses relative to wild-type EBOV. Our results offer insight into the effects of host factors on the evolution of EBOV and highlight the capacity of the virus to rapidly develop potentially adaptive mutations in diverse hosts. Given the ongoing severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic, and the likely origin of this virus in bats, expanding our understanding of the evolution of viruses in their bat hosts is of particular relevance at this time.

RESULTS

Experimental Evolution through Serial Passaging of EBOV in Human and Bat Cells

The cell line EpoNi/22.1, derived from renal epithelia of an adult *Epomops buettikoferi*, was selected as the bat cell line used for

trends (e.g., population growth rate, bottlenecks), affect observed evolutionary rates. Furthermore, these estimates of EBOV evolutionary rates are derived from consensus sequences obtained from human cases and do not reflect evolution in the natural reservoir of the virus. Although the effects of host-specific conditions on the observed mutation rate of EBOV are unknown and may or may not differ between reservoir and non-reservoir hosts, the factors that dictate evolutionary rate during circulation (i.e., positive/negative selection, genetic drift) likely vary (Holmes et al., 2016). Experimental data demonstrate that the animal passage history of EBOV influences its infectivity and virulence during subsequent infection of a new host species, and a similar effect is presumed to occur in natural settings (Gale et al., 2016).

The 2013–2016 West African EBOV epidemic generated an unprecedented abundance of sequencing data. Several fixed putative adaptive mutations were identified. Furthermore, at least two and possibly three of these were under positive selection (Diehl et al., 2016; Dietzel et al., 2017; Urbanowicz et al., 2016). Despite exhibiting increased fitness in cell culture, no obvious difference in pathogenicity from the parental virus was found in mouse and rhesus macaque models of EBOV infection (Marzi et al., 2018).

passaging. This species is a close relative of *Epomops franqueti* (Hoffmann et al., 2013). 293T cells, derived from human embryonic kidney, were used for comparison. These cell lines were selected because of their similar tissue origin and the fact that they replicate the virus to similar titers. The latter is an important consideration for population genetics, as vastly divergent population sizes complicate analysis.

EBOV was rescued from the full-length clone plasmid in 293T cells. Passage 2 (p2) virus was blind-passaged three times in either EpoNi/22.1 or 293T cell lines for initial "adaptation." This step reduces the risk of interference from any extremely high-fitness mutations associated with early passage in each cell line that may either obscure or artificially inflate the fitness of lower frequency mutations. Next, viruses were put through two rounds of terminal dilution. Three clonal isolates were selected from each cell line. Resulting titers were low, and two rounds of amplification in their respective cell lines were required. For experimental passages, monolayers of 293T or EpoNi/22.1 were inoculated at multiplicity of infection (MOI) 0.1 plaque-forming units (PFU)/cell. The first passage in 293T was performed at MOI 0.01 PFU/cell for all replicates, as the titer was low following the amplification passages. This process was repeated for a total of seven experimental passages. Supernatants from each passage were collected, and virus was purified via sucrose gradient for subsequent RNA extraction and sequencing. Figure 1 presents an overview of the experiment. In 293T, titers for clones A and C remained relatively stable throughout the passage series; however, starting in passage 6, a precipitous decline in titer was observed for clone B (Figure S1). No substantial difference was observed in the three clones passaged in EpoNi/22.1.

EBOV Takes Distinct Evolutionary Paths in Human-Derived and Bat-Derived Cell Lines

Viral genomic RNA isolated after each passage was used to prepare libraries for CirSeq (Acevedo and Andino, 2014; Acevedo et al., 2014). First, the data were used to calculate individual mutation frequencies for each possible nucleotide variant (A to G, C to A, U to C, etc.) (Figure 2A). Averaged over all clones and passages, we found that overall mutation frequencies were similar between the EpoNi/22.1 and 293T-derived viruses, with the exception of G-to-A transitions. We found there was a significant increase in the frequency of G-to-A mutations (with respect to the EBOV genomic strand) in the EpoNi/22.1-derived viral populations relative to 293T (Figure 2A).

To determine if mutation rates within individual viral genes differed between cell lines, we recalculated individual mutation frequencies, treating each open reading frame (ORF) as an independent region. Frequencies were consistent, indicating no gross differences in the spontaneous RdRP mutation rate due to genomic position (Figure 2B). This also held true for the elevated frequency of G-to-A substitutions observed in EpoNi/22.1-derived viral genomes, with all examined regions exhibiting a similar pattern relative to 293T-derived virus (Figure 2B).

Clear differences in variant frequencies at the final passage highlight the distinct evolutionary paths of the 293T-passaged and EpoNi/22.1-passaged populations (Figures S2A and S2B). Comparing Shannon entropy over time, we found a high degree of homogeneity in passage 1, with increasing heterogeneity over

the course of passaging (Figure S2C). Overall, the average genomic Shannon entropy estimated in the EpoNi/22.1-passaged replicates was moderately higher than in their 293T-passaged counterparts (Figure S2C).

EBOV Populations Passaged in Bat-Derived Cells Exhibit a "Spike" of High-Frequency Mutations Consistent with ADAR Activity in the GC/MLD Region of the GP Protein

An intriguing pattern of mutations arose during passaging of EBOV in EpoNi/22.1 cells. Two of the three EpoNi/22.1-passaged viral populations acquired localized peaks of high-frequency variants over the course of passaging (Figure 3A). These mutations arose primarily after passage 4 and consisted almost entirely of adenosine (A)-to-guanine (G) substitutions (genomic sense). These peaks of mutations were localized within the region spanning the GC and MLD regions of GP (Figure 3B). An increased frequency of A-to-G substitutions in these regions was also detected in the other passage series (including the 293T-derived populations), albeit at a lower frequency.

As a result of the relatively small size of the region in which they were located, we were able to determine if multiple mutations appeared on a single genome. In two of the three replicates passaged in EpoNi/22.1, the average number of A-to-G substitutions (genomic strand) per read (i.e., genome) increased dramatically over the course of passaging (Figure 3C, solid lines), while U-to-C substitutions did not (Figure 3C, dashed lines). EBOV therefore accumulates these A-to-G mutations on the same genomes without apparent detriment. Interestingly, the passage EpoNi/22.1 A exhibited a higher average number of mutations per read than the two other replicates (Figure 3D). Such a high level of mutational robustness is supported by previous studies in other viruses (Lauring et al., 2013).

A compelling explanation for this phenomenon is editing activity associated with the ADAR family of RNA editing enzymes. (Walkley and Li, 2017). ADARs edit double-stranded RNA (dsRNA) by creating adenosine-to-inosine mutations, ultimately resulting in A-to-G substitutions. ADARs have been implicated in editing the genomes of a number of viruses (Cattaneo et al., 1988; Gélinas et al., 2011; Khrustalev et al., 2017; Piontkivska et al., 2017; Samuel, 2012), including EBOV (Dudas et al., 2017; Park et al., 2015; Shabman et al., 2014; Tong et al., 2015; Whitmer et al., 2018). We investigated whether any ADAR motifs were enriched in the highest frequency variants in this region. Examination of the 10 nucleotides surrounding the most frequent A-to-G (genomic strand) variants revealed a motif matching that expected of ADAR editing (5′-[U/A/C] AG/U-3′) (Figures 4A and 4B) (Eggington et al., 2011). Lending further support to the hypothesis of ADAR editing in viral populations derived from EpoNi/22.1, we found that EpoNi/22.1 cells express approximately 12-fold more ADAR1 mRNA than 293T cells. EBOV infection did not significantly increase ADAR1 expression in either cell line (Figure 4C). However, it appears that this enhanced expression is either a feature of *Epomops* bats or their subfamily (Epomophorinae) (Figures 4C and 4D).

In summary, we have found that during our passaging experiments in bat cells, a region encompassing parts of the GC and MLD of EBOV GP undergoes hypermutation in the form of a
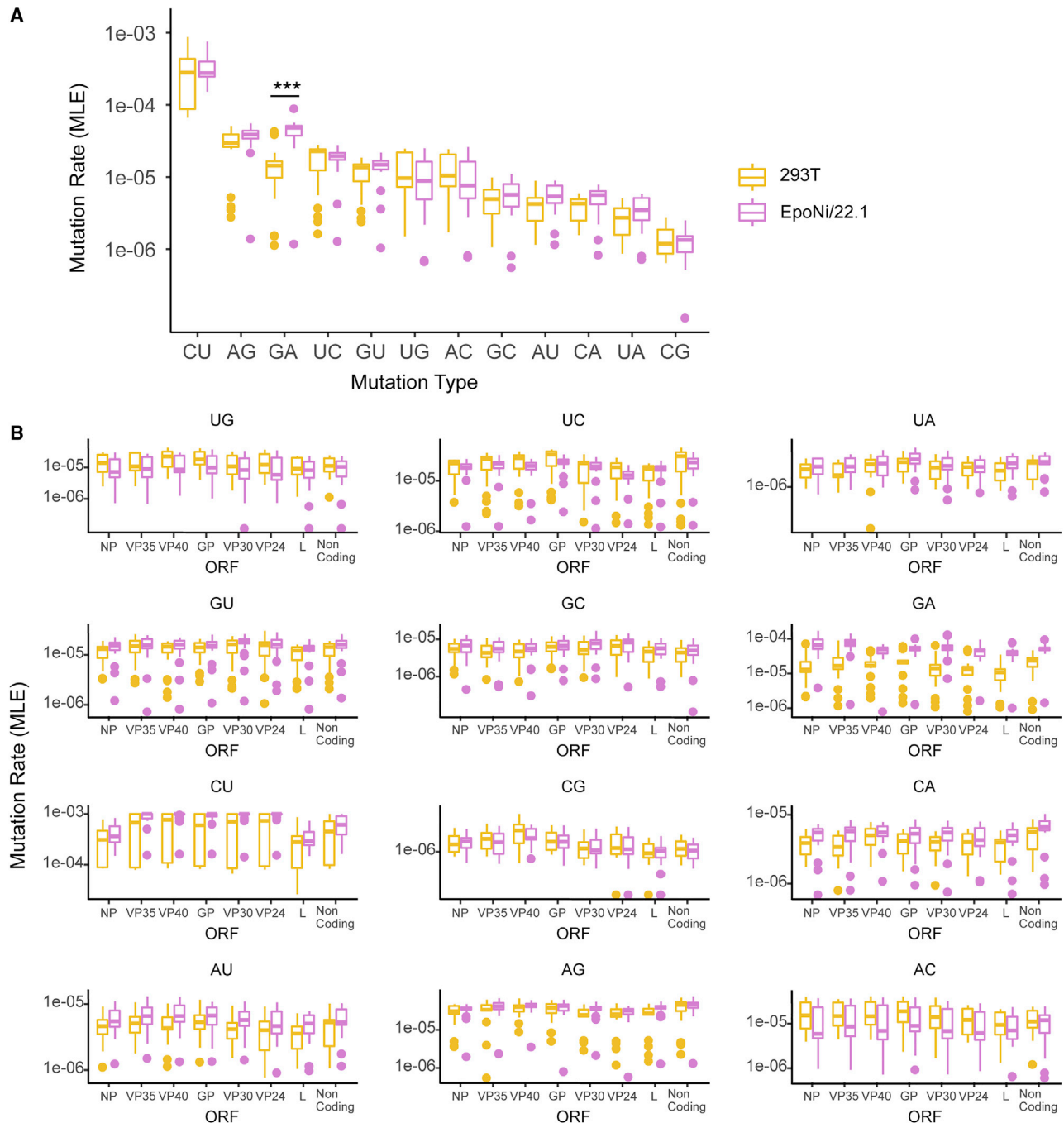
Figure 2. Population-Level Characterization of Passaged EBOV Populations

(A) Mutation frequencies for each type of nucleotide substitution. Each boxplot represents 21 data points (three clones over seven passages). ***Corrected p value $< 10^{-4}$. Error bars denote ±SD.

(B) Identical to (A), but mutation frequencies were calculated on a per ORF basis rather than across the entire genome. G-to-A mutations are significantly elevated in EpoNi/22.1 cells across all ORFs (all corrected p values $< 10^{-3}$). "Non-coding" represents all non-coding regions as a single entity. Error bars denote ±SD.

drastic increase in the rate of A-to-G mutations. These mutations are consistent with the described editing activity of ADAR, an isoform of which was found to be expressed in significantly greater quantities in the bat species or cell line used relative to human cells.

Human and Bat Cell Passage-Produced Viruses Have Distinct Population Structures

In addition to the "spike" of mutations in GP found only in EpoNi/22.1-passaged viruses, we identified individual mutations that rose in frequency over the course of passaging. To identify these,
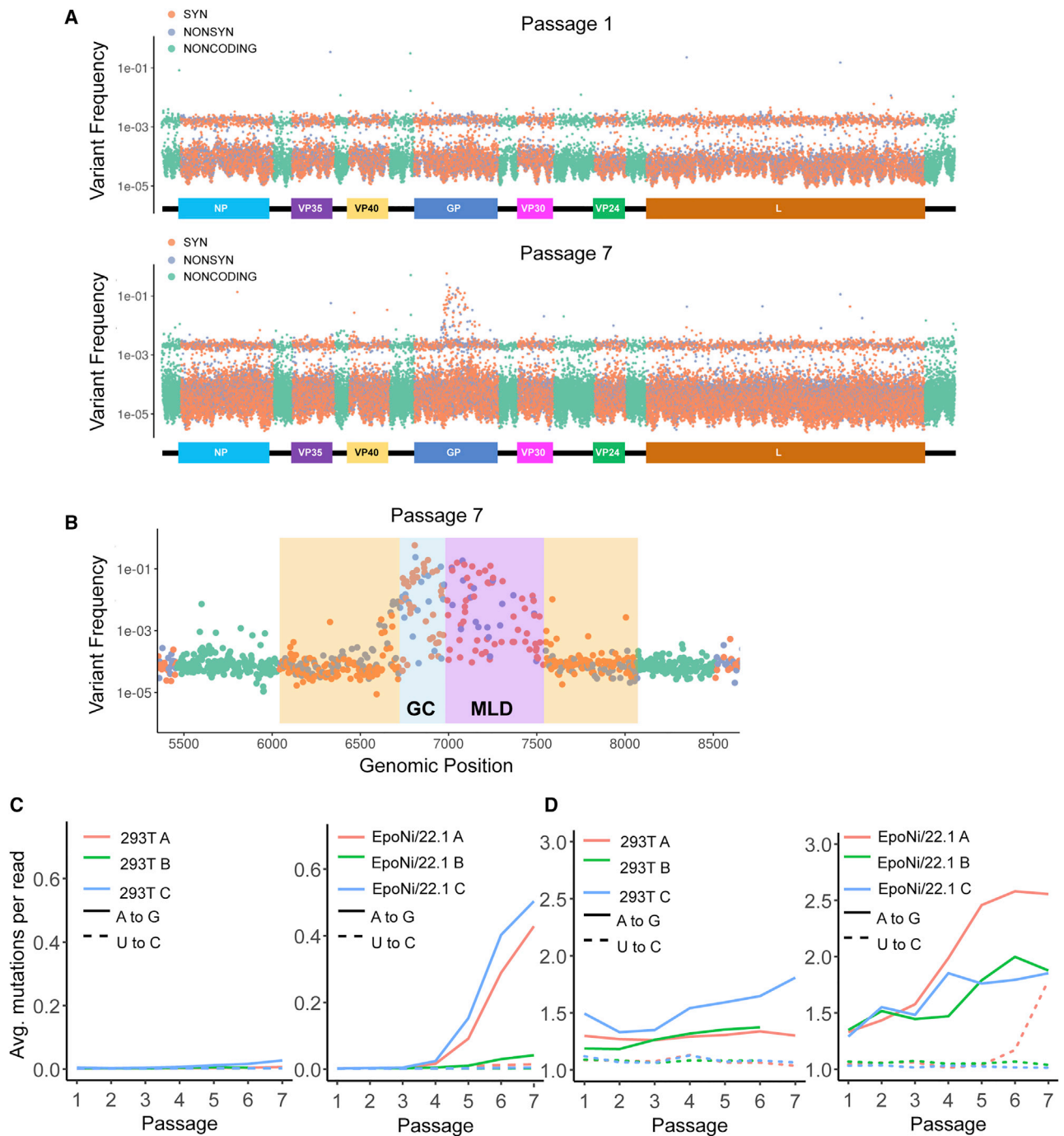
Figure 3. EpoNi-Passaged Virus Accumulates High-Frequency A-to-G Mutations

(A) Variant frequencies plotted against nucleotide position, color-coded by type of amino acid change at passages 1 and 7 of the EpoNi/22.1 "C" population.

(B) Detailed view of the GP coding region, showing only A-to-G (genomic sense) mutations. The majority of mutations in the high-frequency "spike" are A to G and localize primarily within the glycan cap (GC; blue) and mucin-like domain (MLD; purple) of GP (yellow).

(C) Average number of A-to-G (solid lines) and U-to-C (dashed lines) mismatches per read in EpoNi/22.1-derived (left panel) and 293T-derived (right panel) virus genomes. Only reads longer than 80 bp were examined.

(D) Same as (C), but filtered to reads containing at least one mutation of the indicated type.
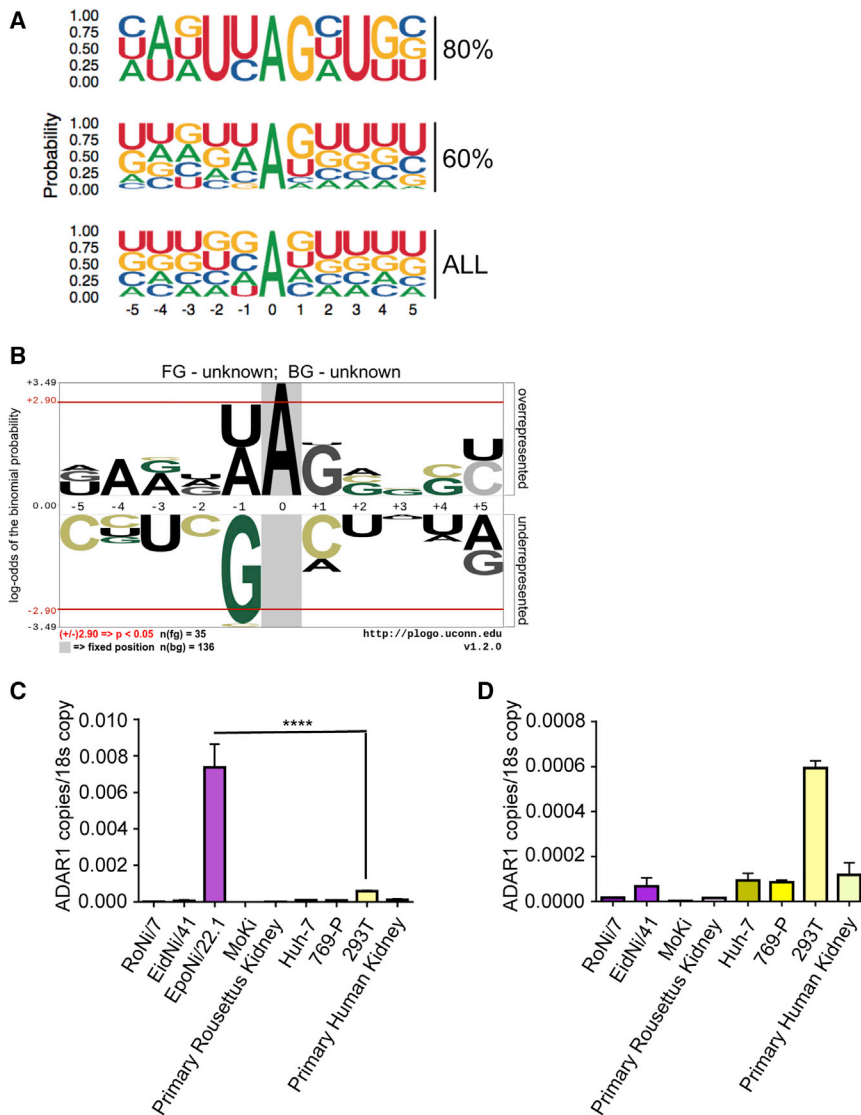
we searched for variants that rose in frequency in at least two of the three EBOV clones passaged in cells derived from each host. Several variants in 293T-derived populations identified were in regions associated with transcriptional regulation. These included mutations in NP, VP30, and the gene-end/transcription termination signal of the VP40 gene (Figure 5; Figure S4). In NP, variants were found within the protein phosphatase 2 (PP2) inter-action domain (Lier et al., 2017), while those in VP30 were near the region of the protein responsible for interaction with NP (Figure S5). Protein modeling revealed that mutations identified in VP30 were predicted to decrease the stability of the protein (Figure S6A). The mutations identified in the VP40 gene-end/transcription termination signal (ATTAAGAAAAAA) (Brauburger et al., 2014) are as follows, with mutated nucleotides underlined: GTTAAGAAAAAA, ATTAGGAAAAAA, ACTAAGAAAAAA, and ATCAAGAAAAAA. These mutations were not generally found to co-occur on the same reads (data not shown). Also identified in 293T passages was a variant cluster within the capping

domain of the L ORF (Figure S5B). Other than the spike of mutations in GP, the only variant cluster identified by visual examination in EpoNi/22.1 was within the methyltransferase domain of the L ORF. Figure S6B illustrates the predicted impacts of the mutations identified in both cell lines on the stability of L polymerase.

Notably, most of the identified mutations did not closely approach fixation in either 293T or EpoNi/22.1 and could be identified only using ultra-deep-sequencing technology, without resorting to a large number of passages. This demonstrates the utility of CirSeq in experimental evolution studies. Variant frequency trajectories for representative mutants that exhibited higher fitness in one cell line are shown in Figure 5. Overall, we found that passaging had host-specific effects on population structure. Therefore, we sought to determine the effects of these differences by characterizing the infection phenotypes of representative mutants.

Human and Bat Cell-Derived Mutants Displayed Cell-Specific Fitness Patterns

A total of six mutants, five of which (VP40 t5885c is excluded) are shown in Figure 5, were selected for characterization. Four mutants were identified in 293T-passaged virus, while two were from EpoNi/22.1-passaged virus. Mutants selected displayed a consistent upward trend in at least two clones during passaging and were generally the most fit within their variant cluster.
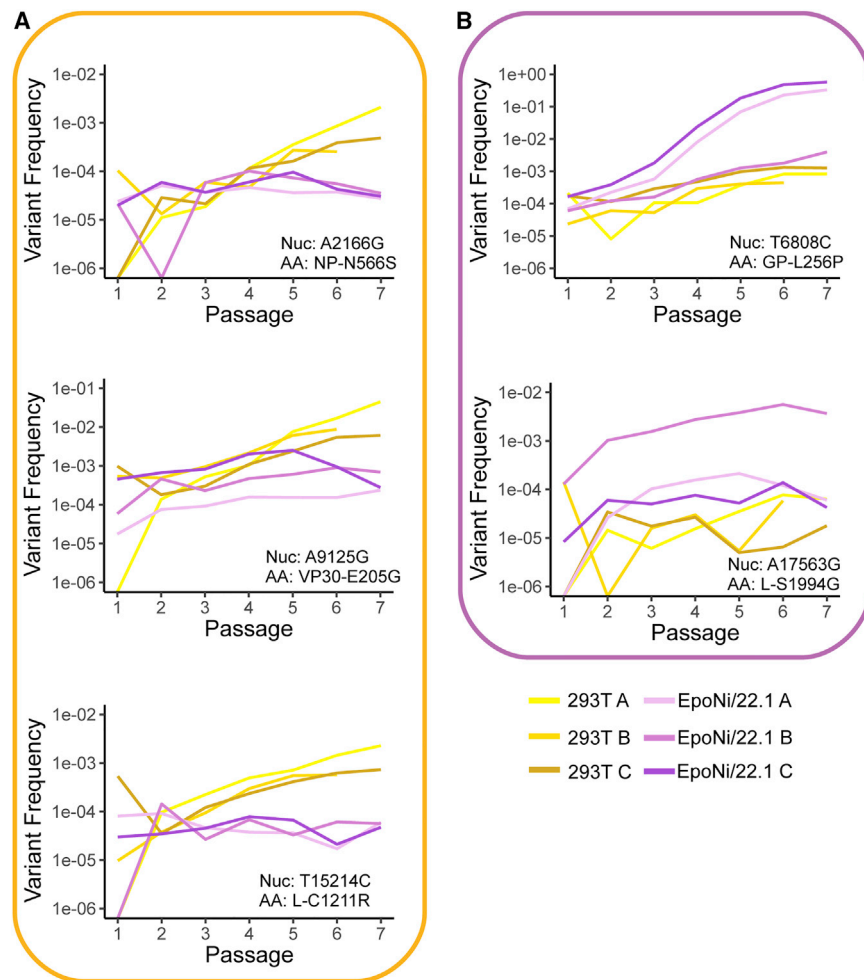
A

B

Figure 5. Frequency Trajectories of Variants with Differential Behavior between Host Cells

(A) Mutations that accumulate faster in 293T-derived populations relative to EpoNi/22.1-derived virus over the course of passaging.

(B) Mutations which accumulate faster in EpoNi/22.1 relative to 293T-derived virus genomes.

shown). Additionally, L S1994G exhibited a large plaque phenotype on Vero E6 (Figure 6B). Finally, a dual luciferase MG assay was used to demonstrate the potential role of noncoding mutations. We found that the mutation identified in the VP40 gene-end/transcription termination signal, t5885c (t5888c on the cDNA clone used to generate the passaged virus, reflected in figures), impaired translation of the second (Renilla luciferase) ORF downstream of the disrupted gene-end signal (Figure 6C).

To better understand the fitness relationships between the mutants and wild-type EBOV, competition assays were performed in both cell lines, as shown in Figure 7A. All mutants were observed to displace the wild-type virus under low-MOI "competition" conditions in both cell lines, with the exception of L C1211R in EpoNi/22.1 cells (Figure 7B), repeating the results of the kinetics assays. However, the kinetics of replacement were variable between viruses and cell lines. The fitness of the 293T-origin mutant L C1211R was very cell line dependent, in contrast to L S1994G, which was detected in EpoNi/22.1 (Figure 7B). Although the EpoNi/22.1-origin GP L256P was more fit than wild-type in both cell lines, its kinetics of displacement were more rapid in EpoNi/22.1 cells, consistent with our replication kinetics results (Figure 7B). The fitness of mutations in polymerase accessory proteins showed little cell line dependency. The only notable trend under high-MOI "complementation" conditions was the slow displacement of NP N566S with VP30 E205G (Figure 7B).

Mutations in ORFs were tested through generation of recombinant viruses using the EBOV full-length clone, while the single untranslated region (UTR) mutant (t5885c) was tested in a VP40/GP bicistronic MG that was developed for this purpose. Replication kinetics assays were performed under multistep conditions in both 293T and EpoNi/22.1 cells. Given the apparent functional relatedness of the mutants, we attempted to construct a double mutant containing the NP and VP30 mutant genes identified in Figure 5. Multiple attempts to rescue this virus failed, suggesting that it is nonviable (data not shown).

Averaged across time points, all mutant viruses had a replicative advantage over the parental wild-type virus in 293T (Figure 6A). In EpoNi/22.1, only GP L256P (identified in EpoNi/22.1) had a meaningful advantage over wild-type (Figure 6A). We also identified cell-specific differences in the infection phenotype of both L polymerase mutants. L C1211R, identified in 293T, exhibited a marked deficiency in EpoNi/22.1 (Figure 6A). The virus also had a small plaque phenotype on Vero E6 (Figure 6B). In comparison, this mutation was neutral to mildly beneficial in 293T. In contrast, single-step kinetics assays performed in EpoNi/22.1 found that L S1994G (identified in EpoNi/22.1) had a significant advantage over wild-type under these conditions (data not

DISCUSSION

The evolution of EBOV in EpoNi/22.1 cells during passaging was remarkably different from that observed in 293T cells. Although divergent evolutionary patterns are not unexpected, the degree and nature of the differences were notable. Although the observed mutation rates were similar in both cell lines, the finding that the rate of G-to-A substitutions was significantly greater in EpoNi/22.1 is particularly important (Figure 2). A potential explanation for this finding is RNA editing of the positive sense complementary RNA (cRNA) by a host factor. C-to-U mutations in the cRNA, such
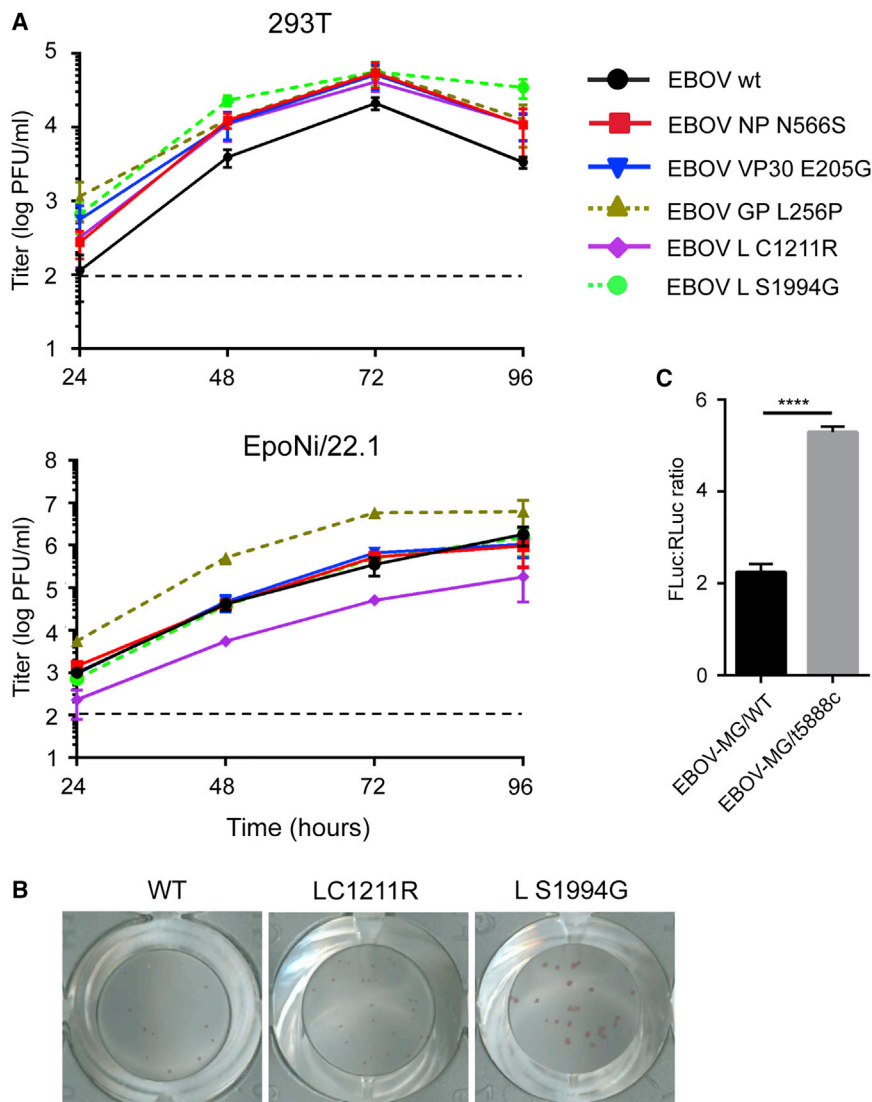
Figure 6. Phenotypic Effects of the Selected Mutations

(A) Multi-step growth kinetics of EBOV wild-type (WT) and mutant viruses in 293T and EpoNi/22.1 cells. Monolayers of 293T or EpoNi/22.1 cells were infected at MOI 0.01 PFU/cell in triplicate. Supernatants collected, then titrated on Vero E6 cells as technical duplicates. 293T-derived and EpoNi/22.1-derived mutants are shown as solid lines and dashed lines, respectively. Limit of detection (2 log PFU/mL) indicated with a dashed line. Error bars denote ±SD.

(B) Immunostain plaque phenotypes.

(C) VP40 gene-end region mutant reduces translation of the second ORF in a bicistronic minigenome. Ratio of firefly luciferase (first ORF) to Renilla luciferase (second ORF) reflects the efficiency of translation of the second ORF. ****$p < 0.0001$ (unpaired t test with Welch's correction). Experiment performed with biological triplicates. No L and no minigenome controls were performed but are not shown. Firefly luciferase activity for these controls was significantly lower than in full plasmid transfections ($p < 0.05$ by one-way ANOVA with Tukey's post hoc test).

activity of a host RNA editing factor. Further investigation demonstrated that this pattern was likely the result of ADAR family RNA editing enzyme activity. ADAR editing of (-)ssRNA (single-stranded RNA) virus genomes is well documented and has been shown to have both proviral and antiviral effects (Gélinas et al., 2011; Samuel, 2012; Suspène et al., 2011). Although 293T-passaged viruses had A-to-G mutations similar to those identified in EpoNi/22.1 cells, the frequency was far lower, suggesting that ADAR activity is elevated in EpoNi/22.1 cells relative to 293T cells.

as those catalyzed by the APOBEC family, would produce G-to-A mutations in the resulting genomic RNA (Harris and Dudley, 2015). Although intriguing, further investigation is required to identify the root cause of this difference and the role host factors may play. Notably, the antiviral effect of APOBEC3 in bats has recently been explored (Hayward et al., 2018).

Evidence for host RNA editing enzyme activity in EpoNi/22.1 cells was found in the GP gene. We observed a spike of high-frequency A-to-G mutations in a region spanning the GC and MLD of GP in EpoNi/22.1-passaged EBOV (Figure 3). These regions are known to be favored targets of the humoral immune response during infection (Flyak et al., 2016), and high-frequency mutations here would be expected in the presence of such strong selective pressure. However, there were no antibodies present during our passaging, and both coding and non-coding mutations were identified. Additionally, the truly massive number of mutations present, and the rate at which these mutations accumulated after passage 4 in EpoNi/22.1 cells, suggests the

Supporting this hypothesis, we found that EpoNi/22.1 cells produce significantly more ADAR1 mRNA than 293T cells. Our data suggest that epomophorine bats express ADAR1 at a higher level than non-epomophorine bat species, which exhibit ADAR1 expression similar to or slightly lower than equivalent human cell lines. However, we found a great deal of variation in ADAR1 expression among bat cell lines derived from a diverse group of bats, compared with relatively limited diversity in expression among human cell lines. This would seem to imply that there may be considerable species-specific differences in ADAR1 expression. A potential limitation of the study is that we cannot exclude the possibility that the variation in levels of ADAR1 across the bat cell lines used is related to the immortalized nature of these cells or reflects a specific stage in the development of the respective organisms. Given the IFN-inducible nature of ADAR1 (Samuel, 2012), it is possible that variable levels of constitutive IFN expression, as has been previously described (Zhou et al., 2016), may be responsible. Considering the
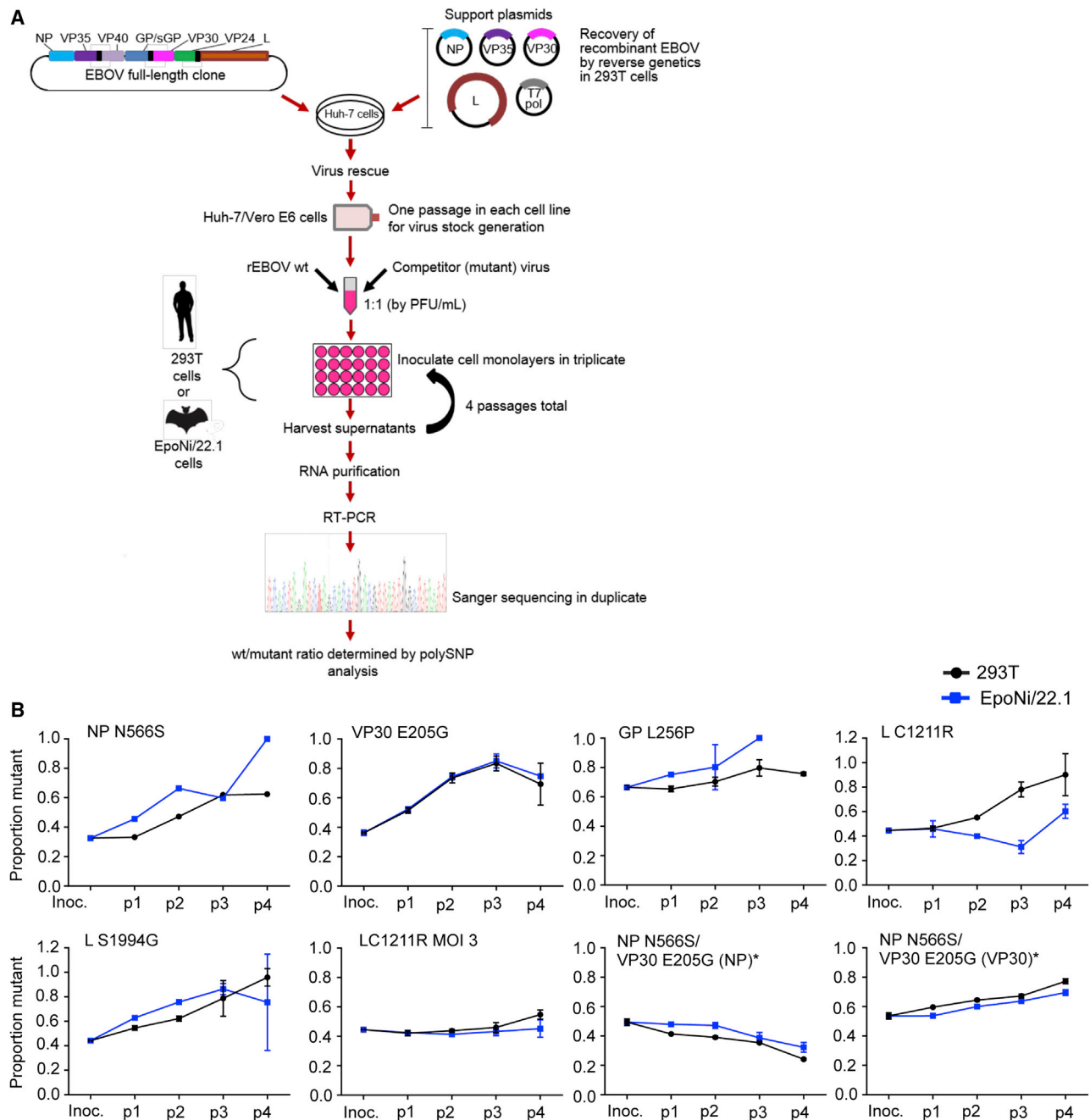
Figure 7. EBOV WT and Mutant Virus Competition and Complementation Assays

(A) Schematic of experimental design.

(B) Data shown are representative of three biological replicates for most competition/complementation assays. Two replicates were analyzed for the following: 293T EBOV WT/NP N566S, EBOV WT/L C1211R p4 competition assays, EpoNi/22.1 EBOV WT/NP N566S and EBOV GP L256P p3 competition assays, and EBOV WT/L C1211R p4 competition assays. One replicate was analyzed for the EpoNi/22.1 EBOV WT/VP30 E205G p4 competition assay. Additionally, because of the inability to generate an amplicon suitable for Sanger sequencing, EpoNi/22.1 EBOV WT/NP N566S and WT/GP L256P p4 competition assays were omitted from the analysis. Sequencing of RT-PCR amplicons was performed with technical duplicates, with the average of the two proportions being used. Error bars denote ± SD. Asterisks denote complementation assays, relative frequency of gene in parentheses displayed.

remarkable diversity of bats (Kunz and Racey, 1998), this would not be unduly surprising. Finally, although it would be preferable to conduct gene silencing experiments to definitively establish the role of ADAR in the hypermutation of GP in EpoNi/22.1 cells, this was not technically feasible due to the lack of a publicly available genome for *Epomops buettikoferi*.

EBOV's MLD and GC are quite flexible and in cell culture appear to be at least partially expendable for GP-pseudotyped vesicular stomatitis virus (Lennemann et al., 2014). However, whether this applies to genuine virus is not well established. We would thus expect editing of these regions to be well tolerated. However, instead of simply being fitness neutral, we found that some of the observed mutations may have been subject to positive selection in EpoNi/22.1cells, where they closely approached fixation (Figure 5B). One mutant reconstituted using the reverse genetics system rapidly displaced wild-type virus in competition assays (Figure 7B). The latter implies that clustering of these mutations on a single genome was not required for increased fitness. Thus, the rapid rise in A-to-G mutations in the EpoNi/22.1-derived populations was likely the result of enzymatic activity and selection.

Evidence of ADAR editing of EBOV genomes has been found in sequences obtained from human cases (Dudas et al., 2017; Park et al., 2015; Whitmer et al., 2018). Specifically, ADAR-like mutations in GP have been reported, although the activity was less specific than what we have observed (Whitmer et al., 2018). Our findings raise the possibility that there has been selective pressure to make GP a favorable target for ADAR. The nucleotide compositions of the MLD and GC show a distinct enrichment for "G" nucleotides, and depletion of "A" nucleotides (Figure S7A) (Khrustalev et al., 2015). This increased frequency of "G" contributes to a uniformly high concentration of the 5′-AG-3′ dinucleotide, part of ADAR's preferred 5′(C/A/U)-AG-3′ target motif (Eggington et al., 2011; Kuttan and Bass, 2012) across the entire GC/MLD region, compared with the other dinucleotides (shown as 3′-GA-5′ in Figure S7B). This specific region is one of the few in the entire genome where 5′-AG-3′ is the most prevalent 5′-AX-3′ dinucleotide. However, the frequency of 5′-GA-3′ (a motif not preferred by ADAR) is also increased in this region. ADAR-driven evolution has been proposed previously for Zika virus and rhabdovirus sigma (Piontkivska et al., 2017). Here, we are reporting evidence for ADAR-driven evolution of portions of the envelope GP that are heavily targeted by the humoral immune response. Taking these facts into consideration, increased susceptibility of this region to ADAR editing may be a strategy to provide an intrinsic means of rapidly generating antibody escape mutants.

In addition to investigating the direct effects of host factors on viral evolution, we also describe changes in population structure that occurred as the result of the virus responding to the replicative environments imposed by the cell lines used. Broadly speaking, purifying selection of EBOV genomes appeared to be a predominant factor in EpoNi/22.1 cells. This is demonstrated by our observation of an increased rate of specific mutations and moderately higher average Shannon entropy in EpoNi/22.1 cells (Figure 2; Figure S3). Thus diversity was higher, but we identified fewer mutations that exhibited positive fitness compared with 293T cells (Figure 5; Figure S2). Moving beyond this global view, interpretation of our data must be conservative. Given the disparities in complexity, *in vitro* evolution cannot always be directly compared with *in vivo* evolution. However, patterns of mutations can be reasonably examined for the purposes of understanding aspects of the more general nature of viral evolution and adaptation in a given species. Therefore, our goal was

to identify regions of the genome that appeared to be responding to the selective pressures imposed by each cell line. In doing this, our focus was on clusters of mutations rather than individual point mutations. This approach has been used previously in tandem with CirSeq in the context of poliovirus (Acevedo et al., 2014).

We identified a number of variant clusters associated with passaging in each cell line. In the ORFs, we identified one cluster in NP proximal to the VP30 binding domain, one in VP30 proximal to the NP binding domain, and one within the capping region of the L gene (Figures S2 and S5). An additional set of mutations was identified in the gene-end signal of VP40. A single representative mutant was selected from each identified cluster. The nature of the NP and VP30 mutants is particularly notable. NP N566S falls within a region reported to be important for interaction with host PP2 (Lier et al., 2017), which is recruited to viral intracytoplasmic inclusion bodies by NP (Lier et al., 2017). PP2 participates in the regulation of EBOV transcription via dephosphorylation of VP30, a requirement for transcription initiation (Martínez et al., 2008; Modrof et al., 2002). VP30 E205G, meanwhile, is proximal to the NP interaction domain of VP30 (Kirchdoerfer et al., 2016) and would likely disrupt an α-helix, significantly disturbing the conformation of the binding domain. Given that NP/VP30 interaction is required for the dynamic phosphorylation of VP30 (Lier et al., 2017; Xu et al., 2017), both of these mutants are predicted to affect EBOV transcription. A double mutant incorporating both VP30 E205G and its equivalent in NP (NP N566S) was nonviable and failed to rescue, implying that the mutants are not complementary but may instead represent an example of convergent evolution. The relative lack of information regarding the structure and function of EBOV L polymerase makes discussion of the potential effects of the capping domain mutant L C1211R difficult, though the mutant and its associated variant cluster may affect the efficiency of mRNA production or on mRNA stability. In contrast, the likely implications of the VP40 gene-end signal mutations are more predictable. Disruption of this highly conserved sequence is almost certain to lead to the production of bicistronic mRNAs, as has been previously described (Brauburger et al., 2014, 2015; Mühlberger, 2007). The second ORF in a bicistronic EBOV mRNA is translated at a drastically reduced frequency, therefore reducing the production of the resulting protein (Brauburger et al., 2015). Our findings with a representative mutant in the EBOV MG system are consistent with this. In EpoNi/22.1 cells, two clusters were identified, the putative ADAR cluster in GP, and another in the methyltransferase domain of L polymerase that likely has similar effects to the L cluster in 293T (Figures S4 and S5).

In competition and complementation assays performed with rescued mutant viruses, we found that most had a fitness advantage over wild-type virus in both cell lines, suggesting that they were genuinely under positive selection during our passage series (Figure 7B). However, there were notable cell line-dependent differences in fitness for specific mutants. These phenotypes are likely the result of differences in the cellular microenvironments of the cell lines and may be worthy of future exploration. The performance of L C1211R was somewhat unexpected. Although selection appeared to be consistently beneficial in all three 293T passage series, it had a very marginal fitness advantage in

293T, the cell line in which it was identified, and was less fit than wild-type in EpoNi/22.1. The displacement of NP N566S by VP30 E205G during complementation assays in both cell lines was also unexpected because the fitness of both viruses relative to wild-type in kinetics assays appeared to be similar. It is likely that the difference is relatively small and that as a result was observable only when the viruses were in direct competition. This hypothesis is bolstered by the low absolute frequency of the mutants in the sequencing data. Our failure to rescue the NP N566S/VP30 E205G double mutant suggests that these mutations are mutually exclusive and that cellular co-infection would not be productive. With complementation therefore impossible, if one virus had even a narrow competitive advantage over the other, it would eventually become the dominant genome.

Taken as a whole, our results validate CirSeq as a predictive tool for the identification of variants and variant clusters associated with increased fitness and adaptive evolution. Moreover, we were able to detect these mutants within a relatively short passage series. In both cell lines, variant clusters in the polymerase are associated with potentially adaptive evolution, but 293T cells produced more adaptive variant clusters. These findings are consistent with studies from the 2014–2016 West African epidemic, which revealed that prolonged passaging in humans induced mutations in the NP, VP30, GP, and L genes (Dietzel et al., 2017; Urbanowicz et al., 2016). We attempted to co-localize select mutations identified in the 2014 West African outbreak (Gire et al., 2014) with mutations exhibiting positive fitness in our passaging experiments (Figures S6C and S6D). The GP1 clamp/base shows an interesting (though non-overlapping) cluster of mutations (in three-dimensional [3D] space) from both the 2014 epidemic and our passaging experiments.

We have identified several key differences in the evolution of EBOV in a human cell line relative to a cell line derived from a close relative of a potential reservoir host. By comparison with the dramatic differences in replicative and fitness environments faced by arboviruses in their host/vector life cycles, the cell lines used in this study are not extraordinarily divergent as both are of mammalian origin. In this light, our identification of a number of meaningful differences in the short-term evolution of the virus in these cell lines is remarkable. We have presented evidence suggesting that RNA editing enzymes play a greater role in the replication and evolution of EBOV in bat cells. As a result of our findings, we propose that ADAR, a host RNA editing enzyme, may have a role in the evolution of the virus in at least one of the cell lines used. Furthermore, we identified regions within the viral genome associated with potentially adaptive evolution resulting from passaging in these cell lines and characterized selected mutations from these regions. Curiously, many of the mutants identified in variant clusters associated with passaging in these cell lines displayed similar, but not identical, fitness in each cell line, suggesting that relatively minor differences in selective pressures could be responsible for the evolutionary divergence we observed. Overall, our findings would suggest that evolution of EBOV in EpoNi/22.1 cells, and potentially by extension in bats, is driven to a significant degree by host factors acting on the genome. By contrast, EBOV evolution in 293T cells appears to be adaptive, with emphasis on regulation of transcription and

transcript stability, as evidenced by variant clusters found within regions of the genome associated with these functions. This pattern fits expectations for a virus that uses bats as a natural reservoir, as evolution in the reservoir host would be drift-driven, while evolution in an incidental host would be more likely to favor positive selection for adaptation (Holmes, 2009; Urbanowicz et al., 2016).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials availability
  - Data and software availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Viruses
- METHOD DETAILS
  - Preparation recombinant wild-type virus stocks for passaging
  - Titration
  - Experimental passages
  - Virus purification
  - Sequencing and processing
  - Determination of population mutation frequency
  - Identification of 'ADAR' motif
  - Average number of mutations per read
  - Fitness estimation of variants
  - Visualization of PDB files and determination of ΔΔG value
  - Calculation of average Shannon's entropy
  - Preparation of mutant viruses
  - Preparation of minigenomes
  - Quantitative analysis of ADAR1 RNA
  - Multistep replication kinetics
  - Competition and complementation assays
  - Minigenome assays
- QUANTIFICATION AND STATISTICAL ASSAYS

REFERENCES

Acevedo, A., and Andino, R. (2014). Library preparation for highly accurate population sequencing of RNA viruses. Nat. Protoc. 9, 1760–1769.

Acevedo, A., Brodsky, L., and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505, 686–690.

Agua-Agum, J., Allegranzi, B., Ariyarajah, A., Aylward, R., Blake, I.M., Barboza, P., Bausch, D., Brennan, R.J., Clement, P., Coffey, P., et al.; WHO Ebola Response Team (2016). After Ebola in West Africa—unpredictable risks, preventable epidemics. N. Engl. J. Med. 375, 587–596.

Andino, R., and Domingo, E. (2015). Viral quasispecies. Virology 479-480, 46–51.

Brauburger, K., Boehmann, Y., Tsuda, Y., Hoenen, T., Olejnik, J., Schümann, M., Ebihara, H., and Mühlberger, E. (2014). Analysis of the highly diverse gene borders in Ebola virus reveals a distinct mechanism of transcriptional regulation. J. Virol. 88, 12558–12571.

Brauburger, K., Boehmann, Y., Krähling, V., and Mühlberger, E. (2015). Transcriptional regulation in Ebola virus: effects of gene border structure and regulatory elements on gene expression and polymerase scanning behavior. J. Virol. 90, 1898–1909.

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.

Coffey, L.L., Beeharry, Y., Bordería, A.V., Blanc, H., and Vignuzzi, M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. Proc. Natl. Acad. Sci. U S A 108, 16038–16043.

Diehl, W.E., Lin, A.E., Grubaugh, N.D., Carvalho, L.M., Kim, K., Kyawe, P.P., McCauley, S.M., Donnard, E., Kucukural, A., McDonel, P., et al. (2016). Ebola virus glycoprotein with increased infectivity dominated the 2013-2016 epidemic. Cell 167, 1088–1098.e6.

Dietzel, E., Schudt, G., Krähling, V., Matrosovich, M., and Becker, S. (2017). Functional characterization of adaptive mutations during the West African Ebola virus outbreak. J. Virol. 91, e01913-16.

Dolan, P.T., Taguwa, S., Rangel, M.A., Acevedo, A., Hagai, T., Andino, R., and Frydman, J. (2020). Principles of dengue virus evolvability derived from genotype-fitness maps in human and mosquito cells. bioRxiv. https://doi.org/10.1101/2020.02.05.936195.

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.

EcoHealth Alliance (2019). EcoHealth Alliance scientists discover the deadly Zaire Ebola virus in West African bat. https://www.ecohealthalliance.org/2019/01/ecohealth-alliance-scientists-discover-the-deadly-zaire-ebola-virus-in-west-african-bat.

Eggington, J.M., Greene, T., and Bass, B.L. (2011). Predicting sites of ADAR editing in double-stranded RNA. Nat. Commun. 2, 319.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzera, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., et al. (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science 288, 339–344.

Flyak, A.I., Shen, X., Murin, C.D., Turner, H.L., David, J.A., Fusco, M.L., Lampley, R., Kose, N., Ilinykh, P.A., Kuzmina, N., et al. (2016). Cross-reactive and potent neutralizing antibody responses in human survivors of natural ebolavirus infection. Cell 164, 392–405.

Gale, P., Simons, R.R., Horigan, V., Snary, E.L., Fooks, A.R., and Drew, T.W. (2016). The challenge of using experimental infectivity data in risk assessment for Ebola virus: why ecology may be important. J. Appl. Microbiol. 120, 17–28.

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

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

Goldstein, T., Anthony, S.J., Gbakima, A., Bird, B.H., Bangura, J., Tremeau-Bravard, A., Belaganahalli, M.N., Wells, H.L., Dhanota, J.K., Liang, E., et al. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. Nat. Microbiol. 3, 1084–1089.

Grubaugh, N.D., Smith, D.R., Brackney, D.E., Bosco-Lauth, A.M., Fauver, J.R., Campbell, C.L., Felix, T.A., Romo, H., Duggal, N.K., Dietrich, E.A., et al. (2015). Experimental evolution of an RNA virus in wild birds: evidence for host-dependent impacts on population structure and competitive fitness. PLoS Pathog. 11, e1004874.

Grubaugh, N.D., Weger-Lucarelli, J., Murrieta, R.A., Fauver, J.R., Garcia-Luna, S.M., Prasad, A.N., Black, W.C., 4th, and Ebel, G.D. (2016). Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. Cell Host Microbe 19, 481–492.

Hall, G.S., and Little, D.P. (2007). Relative quantitation of virus population size in mixed genotype infections using sequencing chromatograms. J. Virol. Methods 146, 22–28.

Harris, R.S., and Dudley, J.P. (2015). APOBECs and virus restriction. Virology 479-480, 131–145.

Hayward, J.A., Tachedjian, M., Cui, J., Cheng, A.Z., Johnson, A., Baker, M.L., Harris, R.S., Wang, L.F., and Tachedjian, G. (2018). Differential evolution of antiretroviral restriction factors in pteropid bats as revealed by APOBEC3 gene complexity. Mol. Biol. Evol. 35, 1626–1637.

Hoffmann, M., Müller, M.A., Drexler, J.F., Glende, J., Erdt, M., Gützkow, T., Losemann, C., Binger, T., Deng, H., Schwegmann-Weßels, C., et al. (2013). Differential sensitivity of bat cells to infection by enveloped RNA viruses: coronaviruses, paramyxoviruses, filoviruses, and influenza viruses. PLoS ONE 8, e72942.

Holmes, E.C. (2009). The Evolution and Emergence of RNA Viruses (Oxford: Oxford University Press).

Holmes, E.C., Dudas, G., Rambaut, A., and Andersen, K.G. (2016). The evolution of Ebola virus: insights from the 2013-2016 epidemic. Nature 538, 193–200.

Ilinykh, P.A., Tigabu, B., Ivanov, A., Ammosova, T., Obukhov, Y., Garron, T., Kumari, N., Kovalskyy, D., Platonov, M.O., Naumchik, V.S., et al. (2014). Role of protein phosphatase 1 in dephosphorylation of Ebola virus VP30 protein and its targeting for the inhibition of viral transcription. J. Biol. Chem. 289, 22723–22738.

Ivanov, A., Ramanathan, P., Parry, C., Ilinykh, P.A., Lin, X., Petukhov, M., Obukhov, Y., Ammosova, T., Amarasinghe, G.K., Bukreyev, A., and Nekhai, S. (2020). Global phosphoproteomic analysis of Ebola virions reveals a novel

role for VP35 phosphorylation-dependent regulation of genome transcription. Cell. Mol. Life Sci. 77, 2579–2603.

Khrustalev, V.V., Barkovsky, E.V., and Khrustaleva, T.A. (2015). Local mutational pressures in genomes of Zaire Ebolavirus and Marburg virus. Adv. Bioinforma. 2015, 678587.

Khrustalev, V.V., Khrustaleva, T.A., Sharma, N., and Giri, R. (2017). Mutational pressure in Zika virus: local ADAR-editing areas associated with pauses in translation and replication. Front. Cell. Infect. Microbiol. 7, 44.

Kirchdoerfer, R.N., Moyer, C.L., Abelson, D.M., and Saphire, E.O. (2016). The Ebola virus VP30-NP interaction is a regulator of viral RNA synthesis. PLoS Pathog. 12, e1005937.

Kuhn, J.H., Amarasinghe, G.K., Basler, C.F., Bavari, S., Bukreyev, A., Chandran, K., Crozier, I., Dolnik, O., Dye, J.M., Formenty, P.B.H., et al.; ICTV Report Consortium (2019). ICTV virus taxonomy profile: Filoviridae. J. Gen. Virol. 100, 911–912.

Kunz, T.H., and Racey, P.A. (1998). Bat Biology and Conservation (Washington, DC: Smithsonian Institution Press).

Kuttan, A., and Bass, B.L. (2012). Mechanistic insights into editing-site specificity of ADARs. Proc. Natl. Acad. Sci. U S A 109, E3295–E3304.

Lauring, A.S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. Nat. Rev. Microbiol. 11, 327–336.

Lennemann, N.J., Rhein, B.A., Ndungo, E., Chandran, K., Qiu, X., and Maury, W. (2014). Comprehensive functional analysis of N-linked glycans on Ebola virus GP1. MBio 5, e00862-13.

Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Délicat, A., Paweska, J.T., Gonzalez, J.P., and Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. Nature 438, 575–576.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Lier, C., Becker, S., and Biedenkopf, N. (2017). Dynamic phosphorylation of Ebola virus VP30 in NP-induced inclusion bodies. Virology 512, 39–47.

Lubaki, N.M., Ilinykh, P., Pietzsch, C., Tigabu, B., Freiberg, A.N., Koup, R.A., and Bukreyev, A. (2013). The lack of maturation of Ebola virus-infected dendritic cells results from the cooperative effect of at least two viral domains. J. Virol. 87, 7471–7485.

Marí Saéz, A., Weiss, S., Nowak, K., Lapeyre, V., Zimmermann, F., Düx, A., Kühl, H.S., Kaba, M., Regnaut, S., Merkel, K., et al. (2015). Investigating the zoonotic origin of the West African Ebola epidemic. EMBO Mol. Med. 7, 17–23.

Martínez, M.J., Biedenkopf, N., Volchkova, V., Hartlieb, B., Alazard-Dany, N., Reynard, O., Becker, S., and Volchkov, V. (2008). Role of Ebola virus VP30 in transcription reinitiation. J. Virol. 82, 12569–12573.

Marzi, A., Chadinah, S., Haddock, E., Feldmann, F., Arndt, N., Martellaro, C., Scott, D.P., Hanley, P.W., Nyenswah, T.G., Sow, S., et al. (2018). Recently identified mutations in the Ebola virus-Makona genome do not alter pathogenicity in animal models. Cell Rep. 23, 1806–1816.

Mbala-Kingebeni, P., Aziza, A., Di Paola, N., Wiley, M.R., Makiala-Mandanda, S., Caviness, K., Pratt, C.B., Ladner, J.T., Kugelman, J.R., Prieto, K., et al. (2019). Medical countermeasures during the 2018 Ebola virus disease outbreak in the North Kivu and Ituri Provinces of the Democratic Republic of the Congo: a rapid genomic assessment. Lancet Infect. Dis 19, 648–657.

Meng, T., and Kwang, J. (2014). Attenuation of human enterovirus 71 high-replication-fidelity variants in AG129 mice. J. Virol. 88, 5803–5815.

Modrof, J., Mühlberger, E., Klenk, H.D., and Becker, S. (2002). Phosphorylation of VP30 impairs ebola virus transcription. J. Biol. Chem. 277, 33099–33104.

Mühlberger, E. (2007). Filovirus replication and transcription. Future Virol. 2, 205–215.

Mühlberger, E., Weik, M., Volchkov, V.E., Klenk, H.D., and Becker, S. (1999). Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. J. Virol. 73, 2333–2342.

Negredo, A., Palacios, G., Vázquez-Morón, S., González, F., Dopazo, H., Molero, F., Juste, J., Quetglas, J., Savji, N., de la Cruz Martínez, M., et al. (2011). Discovery of an ebolavirus-like filovirus in europe. PLoS Pathog. 7, e1002304.

Médecins sans Frontières (2020). New Ebola cases confirmed in DRC days before expected end of outbreak. https://www.msf.org/new-ebola-cases-confirmed-drc.

O'Shea, J.P., Chou, M.F., Quader, S.A., Ryan, J.K., Church, G.M., and Schwartz, D. (2013). pLogo: a probabilistic approach to visualizing sequence motifs. Nat. Methods 10, 1211–1212.

Olival, K.J., and Hayman, D.T. (2014). Filoviruses in bats: current knowledge and future directions. Viruses 6, 1759–1788.

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.

Peck, K.M., and Lauring, A.S. (2018). Complexities of viral mutation rates. J. Virol. 92, e01031-17.

Pfeiffer, J.K., and Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. 1, e11.

Piontkivska, H., Frederick, M., Miyamoto, M.M., and Wayne, M.L. (2017). RNA editing by the host ADAR system affects the molecular evolution of the Zika virus. Ecol. Evol. 7, 4475–4485.

R Core Team (2018). R: a language and environment for statistical computing. https://www.R-project.org/.

RStudio Team (2016). RStudio: integrated development for R. http://www.rstudio.com/.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.

Samuel, C.E. (2012). ADARs: viruses and innate immunity. Curr. Top. Microbiol. Immunol. 353, 163–195.

Schrödinger, LLC (2020). The PyMOL Molecular Graphics System, version 2.0. https://pymol.org/2/.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Res. 33, W382–W388.

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.

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.

Taylor, D.J., Dittmar, K., Ballinger, M.J., and Bruenn, J.A. (2011). Evolutionary maintenance of filovirus-like genes in bat genomes. BMC Evol. Biol. 11, 336.

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

Towner, J.S., Amman, B.R., Sealy, T.K., Carroll, S.A., Comer, J.A., Kemp, A., Swanepoel, R., Paddock, C.D., Balinandi, S., Khristova, M.L., et al. (2009). Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. PLoS Pathog. 5, e1000536.

Tsuda, Y., Hoenen, T., Banadyga, L., Weisend, C., Ricklefs, S.M., Porcella, S.F., and Ebihara, H. (2015). An improved reverse genetics system to overcome cell-type-dependent Ebola virus genome plasticity. J. Infect. Dis. 212 (Suppl 2), S129–S137.

Urbanowicz, R.A., McClure, C.P., Sakuntabhai, A., Sall, A.A., Kobinger, G., Muller, M.A., Holmes, E.C., Rey, F.A., Simon-Loriere, E., and Ball, J.K. (2016). Human adaptation of Ebola virus during the West African outbreak. Cell 167, 1079–1087.e5.

Van Rossum, G., and Drake, F.L., Jr. (1995). Python tutorial, technical report CS-R9526 (Amsterdam: Centrum voor Wiskunde en Informatica [CWI]).

Vignuzzi, M., Stone, J.K., and Andino, R. (2005). Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. Virus Res. *107*, 173–181.

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647.

Walkley, C.R., and Li, J.B. (2017). Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. Genome Biol. *18*, 205.

Whitmer, S.L.M., Ladner, J.T., Wiley, M.R., Patel, K., Dudas, G., Rambaut, A., Sahr, F., Prieto, K., Shepard, S.S., Carmody, E., et al.; Ebola Virus Persistence Study Group (2018). Active Ebola virus replication and heterogeneous evolutionary rates in EVD survivors. Cell Rep. *22*, 1159–1168.

World Health Organization (2020). Ebola situation reports: Democratic Republic of the Congo. https://www.who.int/csr/don/16-April-2020-ebola-drc/en/.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis, 2nd (New York: Springer).

Xu, W., Luthra, P., Wu, C., Batra, J., Leung, D.W., Basler, C.F., and Amarasinghe, G.K. (2017). Ebola virus VP30 and nucleoprotein interactions modulate viral RNA synthesis. Nat. Commun. *8*, 15576.

Yang, X.L., Tan, C.W., Anderson, D.E., Jiang, R.D., Li, B., Zhang, W., Zhu, Y., Lim, X.F., Zhou, P., Liu, X.L., et al. (2019). Characterization of a filovirus (Měnglà virus) from Rousettus bats in China. Nat. Microbiol. *4*, 390–395.

Zhou, P., Tachedjian, M., Wynne, J.W., Boyd, V., Cui, J., Smith, I., Cowled, C., Ng, J.H., Mok, L., Michalski, W.P., et al. (2016). Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. Proc. Natl. Acad. Sci. U S A *113*, 2696–2701.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-EBOV GP pAb	IBT Bioservices	0301-015
Goat Anti-Rabbit IgG-HRP	SouthernBiotech	Cat# 4030-05; RRID:AB_2687483
Bacterial and Virus Strains		
Ebola virus, variant Mayinga	Recombinant from full-length clone	NCBI:txid128952
Deposited Data		
Consensus FASTQ files	This paper	SRA: PRJNA597079
Experimental Models: Cell Lines		
Vero E6	ATCC	C1008
293T	ATCC	CRL-3216
EpoNi/22.1	Dr. Christian Drosten	CVCL_RX73
Huh-7	Lab stock	CVCL_0336
Oligonucleotides		
Primers	IDT	See Tables S1 and S2.
Recombinant DNA		
Ebola virus full length clone	Dr. Stuart Nichol	pEBO
Software and Algorithms		
R version 3.4.4	R Core Team, 2018	https://www.r-project.org/
Rstudio 1.1.453	RStudio Team, 2016	http://www.rstudio.com/
Python 2.7.15	Van Rossum and Drake, 1995	python.org
CirSeq software (with minor updates)	Acevedo et al., 2014	https://andino.ucsf.edu/CirSeq (file with minor update available at https://doi.org/10.17632/42z69y3v35)
Maximum likelihood Estimation of mutation rates (MaximumLikelihoodEstimation_Q20_Zach.R)	This paper Mendeley Data	https://github.com/ashleyacevedo/mutation_rates and Mendeley Data: https://doi.org/10.17632/42z69y3v35
FitnessEstimator	Dolan et al., 2020 (bioRXiv)	http://biorxiv.org/lookup/doi/10.1101/2020.02.05.936195 (bioRXiv)
MultiMatch	Dolan et al., 2020 (bioRXiv)	http://biorxiv.org/lookup/doi/10.1101/2020.02.05.936195 (bioRXiv)
ggplot2	Wickham, 2009	https://ggplot2.tidyverse.org/
pLogo	O'Shea et al., 2013	https://plogo.uconn.edu/
pySam	Li et al., 2009	https://github.com/pysam-developers/pysam
ggseqlogo	Wagih, 2017	https://github.com/omarwagih/ggseqlogo
Pymol	The PyMOL Molecular Graphics System, Version 1.8.4.0 SchrödingerLLC, 2020	https://pymol.org/2/
FoldX	Schymkowitz et al., 2005	http://foldxsuite.crg.eu/
MODELER	Sali and Blundell, 1993	https://salilab.org/modeller/
plot_mutation_rates.R	This Paper, Mendeley Data	Mendeley Data: https://doi.org/10.17632/42z69y3v35
adar_motif_analysis.R	This Paper, Mendeley Data	Mendeley Data: https://doi.org/10.17632/42z69y3v35
mismatchesPerRead_combo_AtoGorTtoC_usingPySam.py	This Paper, Mendeley Data	Mendeley Data: https://doi.org/10.17632/42z69y3v35
visualizeSameRead Distributions_functions_v3_toPowerPoint.R	This Paper, Mendeley Data	Mendeley Data: https://doi.org/10.17632/42z69y3v35

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to Alexander Bukreyev (abukreye@utmb.edu).

Materials availability
All unique reagents generated in this study are available from the Lead Contact with appropriate regulatory clearances and a competed Materials Transfer Agreement.

Data and software availability
The CirSeq consensus fastq files generated during this study are available at the SRA database under accession number PRJNA597079.

Data containing the EBOV genome used for mapping reads ('*Ebola_fixed.fasta*'), variant counts ('*Q20threshold.txt*''), variant frequencies ('*Q20Freqs_SD.txt*'), mutation rates ('*MutationRates_Q20MLE.txt*'), and variant counts/frequencies paired with codon/amino acid information ('*Q20thresholdTranslated.txt*'') for each replicate-passage combination in the paper are available in Mendeley Data (https://doi.org/10.17632/42z69y3v35).

File descriptions are as follows.

Ebola_fixed.fasta: FASTA file of EBOV genome used in the analysis.

Q20threshold.txt: One of the direct outputs of the CirSeq software showing variant counts at each genome position. Column 1 is genome position, column 2 is the reference base, columns 3,4,5, and 6 are the counts for A,C,G, and T/U respectively at that position. In some cases, deposited files are the combined Q20threshold.txt files from multiple sequencing runs.

Q20Freqs_SD.txt: Produced directly from Q20threshold.txt and contains four lines for every position in the genome, each representing a possible variant at that position (including the reference base itself). Column 1 is the genomic position, column 2 is the reference base at the position, column 3 is the potential variant base, column 4 is the counts of the indicated variant base, column 5 is the total counts at the indicated position, column 6 is the frequency of the indicated variant base, and column 7 is an estimate of the standard deviation associated with the given variant.

MutationRates_Q20MLE.txt: Output of 'MaximumLikelihoodEstimation_Q20_Zach.R'. Column 1 is the mutation type, column 2 is the maximum likelihood estimate of the mutation rate, and column 3 is the estimate of standard error.

Q20thresholdTranslated.txt: Provides variant frequency information, but in the context of the protein/codon. Column 1 is the genomic position, column 2 is the amino acid position (NA if non-coding), column 3 is the position within the codon (1,2, or 3), column 4 is the reference codon (single base if non-coding), column 5 is the reference amino acid (NA if non-coding), column 6 is the variant codon (single base if non-coding), column 7 is the variant amino acid (NA if non-coding), column 8 is the counts of the indicated variant, column 9 is the total counts/coverage at the genomic position, column 10 is the frequency of the given variant, column 11 indicates whether the current position is synonymous, nonsynonymous, or noncoding, column 12 indicates the affected protein (or if intergenic).

Code is provided which contributed to Figures 1A, 3C, 4A, 4B, and S3 in Mendeley Data (https://doi.org/10.17632/42z69y3v35).

293T_A_B_C_p1-7.rds/EpoNi_A_B_C_p1-7.rds: Processed fitness information as taken from the output of the FitnessEstimator based on all 7 viral passages for each replicate. These files provide fitness estimates for each possible variant in the genome, upper and lower boundaries of the estimate, and the number of passages in which adequate coverage was detected. Information is provided for all three replicates in each passage series.

Code descriptions are as follows.

preprocessing_3.py: Contains a minor fix to the preprocessing_3.py script included with the CirSeq package.

MaximumLikelihoodEstimation_Q20_Zach.R: Reads in a 'Q20threshold.txt' and produces a maximum likelihood estimate of the overall mutation rate for each possible variant (AtoC, GtoA, etc…).

plot_mutation_rates.R: Plots the organized output of 'MaximumLikelihoodEstimation_Q20_Zach.R' into boxplots separated by host cell line.

adar_motif_analysis.R: Uses 'Q20Freqs_SD.txt' files as input to generate mutations in the specified quantile and plots them using ggseqlogo. This also generates the output for input to pLogo.

mismatchesPerRead_combo_AtoGorTtoC_usingPySam.py: Uses the mapped output (after sorting and converting to bam format) of the CirSeq pipeline to identify TotC or AtoG which occur on the same read. Note the direct output of this script is with respect to the EBOV coding/+ strand.

visualizeSameReadDistributions_functions_v3_toPowerPoint.R: Visualize the output of 'mismatchesPerRead_combo_AtoGorTtoC_usingPySam.py' to visualize . Note the direct output of this script is with respect to the EBOV coding/+ strand. Graphs were altered for the manuscript to be in reference to the EBOV genomic/- strand.

shannonEntropy_avgByPassage_toPowerPoint.R: Uses 'Q20Freqs_SD.txt' files as input to calculate the average Shannon's entropy across the genome per replicate-passage combination.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines
EpoNi/22.1 bat cells were provided by Dr. Christian Drosten as a gift. This cell line has been described previously (Hoffmann et al., 2013). Cells were maintained in DMEM with F12 and GlutaMAX supplements, plus 10% fetal bovine serum (FBS) and 0.1% gentamicin. 293T cells were maintained in DMEM supplemented with 10% FBS and 0.1% gentamicin. Vero E6 cells were maintained in MEM supplemented with 10%5 FBS and 0.1% gentamicin.

Viruses
All viruses used in this study were recombinants based upon a full length clone plasmid provided as a gift by Drs. John Towner and Stuart Nichol (CDC). All work with infectious viruses was performed under BSL-4 conditions by trained personnel in the facilities of Galveston National Laboratory.

METHOD DETAILS

Preparation recombinant wild-type virus stocks for passaging
Rescue of EBOV for the initial passaging experiment was performed as described in Lubaki et al. (2013), using a modified version of the EBOV full length clone provided as a gift by Drs. John Towner and Stuart Nichol (CDC). To generate this clone, the eGFP transgene was excised via restriction digest with BSiWI (New England Biolabs/NEB), following the manufacturer's protocol. The plasmid was re-ligated using T4 DNA ligase (NEB), again following the manufacturer's protocol. The EBOV NP, VP35, L, VP30, and T7 polymerase support plasmids were provided by Dr. Yoshihiro Kawaoka (University of Wisconsin). Following initial rescue, The input virus stock (passage 2) was blind passaged blindly three times ("adaptation passages") in either 293T or EpoNi/22.1 cells, followed by two sequential rounds of terminal dilution in the respective cell lines, from which three clonal virus populations were selected. Isolated viruses were amplified by two blind passages in their respective cell lines to generate viruses with sufficient titers for experimental passages.

Titration
Titration was performed by inoculating confluent monolayers of Vero E6 cells with serially diluted virus allowing the virus to adsorb for 1 hour at 37°C, 5% CO2. Following adsorption, a 0.5% methylcellulose, 2% FBS MEM overlay with 0.1% gentamicin was added and the cells were incubated for 5 days at 37°C, 5% CO_2. Plaques were visualized by plaque immunostaining using an anti-GP polyclonal primary antibody (IBT Bioservices).

Experimental passages
For experimental passages, confluent monolayers of 293T or EpoNi/22.1 were inoculated at MOI 0.1 PFU/cell, except for the first passage in 293T, which was performed at MOI 0.01 PFU per cell due to low titers following the amplification passages. Cells were incubated at 37°C/5% CO_2 for 5 days, after which the supernatants were collected, clarified by centrifugation at 2,000 g, and frozen at −80°C prior to titration and purification. This process was repeated for an additional six experimental passages, at MOI 0.1 PFU/cell.

Virus purification
Viruses were purified for RNA extraction by sucrose gradient ultracentrifugation. Supernatants were layered over 25% sucrose (w/v, diluted in 1X STE buffer), and centrifuged at 175,000x*g* for 2 hours at 4°C. Pelleted virus was resuspended in 0.5 mL of STE buffer and sonicated in a water bath (amplitude 95 Hz) for 30 s. Sonicated samples were layered over a 20%–60% sucrose gradient, topped with 1X STE buffer, and centrifuged at 207,000x*g* for 90 minutes at 4°C. The virus band at the sucrose cushion was collected with a pipette, diluted with 1X STE, and then centrifuged at 207,000x*g* for 1 hour at 4°C. Purified virus pellets were resuspended in 100 μL of 1X STE buffer prior to inactivation in 1 mL of Trizol reagent for removal from the BSL-4 and subsequent RNA extraction following the manufacturer's recommended protocol. For 293T cells, host cell rRNA contamination necessitated removal using the GeneRead rRNA Depletion Kit (QIAGEN).

Sequencing and processing

Libraries for Circular Sequencing (CirSeq) were generated as described previously (Acevedo and Andino, 2014). 300 cycle, single end reads were generated on an Illumina HiSeq 2500 or HiSeq 4000. Resulting fastq files were analyzed as in Acevedo et al. (2014). A small number of additional counts was recovered using the MultiMatch algorithm (Dolan et al., 2020). Variant count files from multiple rounds of sequencing and CirSeq processing were combined to obtain final datasets for analysis. Average coverage per base ranged from 94,461 to 509,722 across all EpoNi/22.1 sequenced libraries. For the 293T libraries, values ranged from 80,797 to 296,930. As a technical note, we did observe some differences in mutation frequency of certain nucleotide substitutions depending on the sequencer used (Illumina HiSeq 2500 versus 4000). The HiSeq 2500 tended to exhibit lower rates of mutation for UG, AC, and CG (Figure S3). However, the differences were not particularly remarkable, and would not be expected to have any meaningful impact on our findings.

Determination of population mutation frequency

A maximum likelihood estimation was used to determine individual mutation frequencies for each nucleotide variant type (A to C, G to A, etc.). Only genomic positions with coverage greater than 100,000 were factored in to the calculation. Mann-Whitney U (stats::wilcox.test() in R) tests were used to assess significance between mutation frequencies of a given variant type. Significance testing for each variant type was performed between the 21 data points (7 passages x 3 replicates) from EpoNi/22.1- and 293T-derived viral populations. Note that 293T B passage 7 was excluded from this analysis due to concerns of potential contamination from another passage.

Identification of 'ADAR' motif

The highest frequency A to G (genomic strand)/T to C (coding strand) mutations centered around GP's mucin-like domain and glycan cap were identified for each replicate (the specific region analyzed was from coding strand nucleotide 6,723 to 7,540 of the EBOV clone used). Positions containing variant frequencies at or above the indicated quantile in all three replicates were used for motif analysis. For example, a given mutation needed to be at or above the 0.8 quantile in passage 7 of EpoNi A, B, and C to be included. Each sequence consisted of the position of interest and its surrounding 10 nucleotides (5 upstream and 5 downstream; 11 nucleotides total). Sequence logos were created using ggSeqLogo. Variants were only considered if their coverage was greater than $3 \times 1/$ (variant frequency).

Average number of mutations per read

The Python package 'pySam' was used to parse a single representative SAM (sorted and converted to BAM) file for each passage output by the CirSeq pipeline. These SAM files represent the mapped consensus read sequences resulting from comparing the head-to-tail repeats generated during the CirSeq workflow. The number of T to C (or A to G) mutations per consensus read was determined, and the average number of each type of mutation per read was determined over the course of passaging for each replicate (including or excluding reads with no mutations). Only reads 80 nt or longer and base calls with a quality score $>= 20$ were used. Note that 293T B passage 7 was excluded from this analysis due to concerns of potential contamination from another passage.

Fitness estimation of variants

Fitness values were calculated using a version of FitnessEstimator (Dolan et al., 2020), using a window size of 6 passages and a bottleneck of 10^6. Significant fitness variants for a given cell line were variants exhibiting beneficial fitness (wrel.ciLower [minimum fitness value in 95% confidence interval] > 1) or deleterious (wrel.ciUpper [maximum fitness value in 95% confidence interval] < 1) in at least two of three clones. Additionally, it was generally required that at least 5 of the 7 passages had high enough coverage at the position of interest to support the calculated frequency (binomial value in FitnessEstimator). The average of these fitness values were used to compare variant fitness between cell lines.

Visualization of PDB files and determination of ΔΔG value

PDB files were visualized using PyMol. A previously published structure of VP30 (5T3T) was used for visualization and stability testing. A structural prediction of the EBOV L protein was constructed using MODELER, using VSV L (5a22) as a template. ΔΔG values were estimated using FoldX. Each PDB file was first repaired (FoldX command = RepairPDB), then a model was built containing the mutation of interest (command = BuildModel). Figure S6 uses structures of GP (PDB structure 3csy) and NP (model from Ivanov et al., 2020).

Calculation of average Shannon's entropy

Entropy was calculated for each nucleotide position in the EBOV genome (at each passage for each replicate). Shannon's entropy for an individual nucleotide was calculated as

$$\text{Entropy}_{single\ nucleotide} = -\Sigma(f * \log_4(f))$$

where 'f' is the frequency (i.e., probability) of each possible nucleotide at that position. The average of this value across the genome was calculated for each replicate at each passage, then plotted. C to U (genomic strand) variants were excluded from the calculations. Effect size (Cohen's d) in the region of the glycan cap and mucin-like domain (defined as nucleotide positions 6723 to 7540), was determined using the cohen.d function from the 'effsize' R package. Distribution of

Shannon entropy per base in all 293T clones was compared to the distribution of Shannon entropies per base for all EpoNi/22.1 clones at each passage. Only positions with coverage greater than 100,000 were evaluated in any calculation of Shannon entropy. This resulted in the comparison of 1,861 positions for EpoNi-derived virus to 2,152 positions for 293T-derived populations across all replicates at passage 6.

Preparation of mutant viruses

Mutant viruses were prepared using either site-directed mutagenesis (NP N566S, VP30 E205G, L C1211R, L S1994G), or by ligating in a synthetic DNA construct (GP L256P). Site-directed mutagenesis was performed using the Q5 Site-directed mutagenesis kit from NEB. Primers were designed using NEB's online tool. As the annealing temperatures suggested by this tool were not successful, an annealing temperature gradient was performed to generate mutant plasmids. Other than this deviation, the manufacturer's protocol was adhered to. For the GP mutant, double-stranded DNA fragments (gBlocks, Integrated DNA Technologies) were first subcloned into the SalI and BbsI sites of a pUC19 construct containing the portion of the EBOV FLC between SalI and SacI. The SalI/SacI fragment of this new construct was digested out and inserted via restriction cloning between the SalI and SacI sites of the FLC plasmid. Viruses were rescued following a modified version of the protocol described by Tsuda et al. (2015). Briefly, 90% confluent 6-well plates of Huh-7 cells in standard maintenance media were transfected with 1 μg pCEZ-NP, 0.5 μg pCEZ-VP35, 0.3 μg pCEZ-VP30, 2 μg pCEZ-L-co, 1 μg PLASMID-T7, and 1 μg of the appropriate FLC plasmid. Transfection complexes were prepared using transIT-LT1 (Mirus), with a ratio of 2 μL of transfection reagent per microgram of plasmid DNA. The next day, media was replaced with fresh DMEM high glucose with 2% FBS and 0.1% gentamicin. Five days post-transfection, supernatants were pooled and adsorbed onto T75 flasks of Huh-7 overnight, with fresh media added the next day. Five days post infection, viruses were collected. To produce stocks of sufficient titer for experiments, viruses were passaged one time on Vero E6 cells infected at MOI 0.1 PFU/cell. Stocks were titrated by immunostaining in 96 well plates.

Preparation of minigenomes

The bicistronic MG was prepared in the Bukreyev lab from a monocistronic MG provided by Dr. Elke Mühlberger (Boston University) (Mühlberger et al., 1999). This MG consists of the 3′ genomic leader, plus the NP 5′ UTR controlling transcription of a firefly luciferase ORF, followed by the VP40-GP gene junction region, including the GP 5′ UTR, which controls transcription of a Renilla luciferase ORF. The Renilla luciferase ORF is followed by the L gene 3′ UTR and the 5′ genomic trailer. Rescue and MG support plasmids (excepting the codon-optimized L plasmid) were as described above. The codon optimized L polymerase plasmid was synthesized by Genescript.

Quantitative analysis of ADAR1 RNA

Actively expanding cell monolayers were lysed in Trizol following manufacturer's protocol. Following RNA extraction, cDNAs were prepared using the iScript cDNA synthesis kit (BioRad) using 20 ng of total RNA. 1 ng of cDNA was used in subsequent qPCR reactions, performed with the iTaq universal SYBR Green Master Mix kit (BioRad). Primers and standards for absolute quantitation were obtained from Integrated DNA Technologies (IDT). Bat ADAR1 and 18 s rRNA primers were designed using the Pteropus vampyrus genome due to the lack of published Epomops genomes for certain species used. Primers and standards used and shown in Table S1. Primers had no more than one centrally placed mismatch relative to their expected targets, where known. qPCR was performed on a QuantStudio 6 thermal cycler. ADAR1 copy number was normalized to 18 s rRNA copy number. Significance was tested using a 1-way ANOVA with a Tukey's post hoc test for multiple comparisons.

Multistep replication kinetics

Assays were performed for all mutants on both EpoNi/22.1 and 293T. Samples were collected at 24, 48, 72, and 96 hours post infection. For each time point, three wells of a 24 well plate were infected at MOI 0.01 PFU/cell, and the virus was allowed to adsorb for 30 minutes, after which wells were washed twice with PBS before fresh media was added. Time course samples were titrated by immunostaining in 96 well plates. Significance was tested using one-way ANOVA, with a Dunnett's post hoc correction.

Competition and complementation assays

Assays were performed by mixing selected mutant viruses 1:1 by immunostain titer. For competition assays, 293T or EpoNi/22.1 cells in 12 well plates were infected for passage 1 at MOI 0.1 PFU/cell in triplicate, otherwise following the infection protocol as described for kinetics assays. Passage 1 infections for complementation assays were performed at MOI 3 PFU/cell. When significant CPE was observed (3 dpi for complementation assays, 4 dpi for competition assays), supernatants were collected. For all subsequent passages, cells for competition assays were infected with a 1:100 dilution of the supernatant from the previous passage. For complementation assays, EpoNi/22.1 cells were infected with ¼ of the supernatant from the previous passage, while 293T cells were infected with ½ of the supernatant from the previous passage, to account for the larger number of 293T cells per well and ensure an MOI greater than 1 PFU/cell. Samples of inocula (in triplicate), and supernatants were removed from BSL-4 in Trizol (Thermo-Fisher) for column-based RNA extraction (Zymo Direct-Zol RNA micro-prep). Figure 7A presents the experimental design in schematic form. RT-PCR amplicons were generated (QIAGEN One Step RT-PCR kit) using the primers provided in Table S2, and were Sanger sequenced with technical duplicates following enzymatic PCR cleanup (Genewiz). Analysis of sequence data was performed

via poly-SNP (Hall and Little, 2007), using the area under the curve method to determine the relative proportion of each virus within the sequenced population. Significance was tested with 2-way ANOVA, with a Bonferroni post hoc correction.

Minigenome assays

MG transfections were performed in triplicate in 6 well plates as previously described (Ilinykh et al., 2014). A codon-optimized L polymerase plasmid was used. Control transfections omitting the L polymerase plasmid were performed with both the wild-type and mutant bicistronic MGs. Dual-luciferase assays were performed to assess the efficiency of translation of the Renilla luciferase ORF relative to the firefly luciferase open reading frame by taking the FFL:RL signal ratio. Comparison of the efficiency of translation from the wild-type MG to translation from the mutant MG was determined by dividing the mutant ratio by the wild-type ratio and taking the reciprocal. This yields a value representative of the loss of efficiency resulting from the mutation. Data presented are representative of three independent experiments. Significance was tested using one-way ANOVA, with a Dunnett's post hoc correction.

QUANTIFICATION AND STATISTICAL ASSAYS

Additional details for specific assays and tests can be found in the relevant Method Details sections and figure legends.

Statistical differences in mutation rate (Figure 1) were determined using the Mann-Whitney U (stats::wilcox.test() in R) followed by multiple comparison testing using the Bonferroni method ('*p.adjust(method = "bonferroni")*' in R). Significance testing for each variant type was performed between the 21 data points (7 passages x 3 replicates) from EpoNi/22.1- and 293T-derived viral populations. Note that 293T B passage 7 was excluded from this analysis due to concerns of potential contamination from another passage.

Effect size of the difference in Shannon's entropy between EpoNi- and 293T-derived viral populations was determined using the effsize::cohen.d() function in R. Effect size was only compared for the individual nucleotides from position 6723 to 7540 (with respect to the reference genome used in this manuscript).
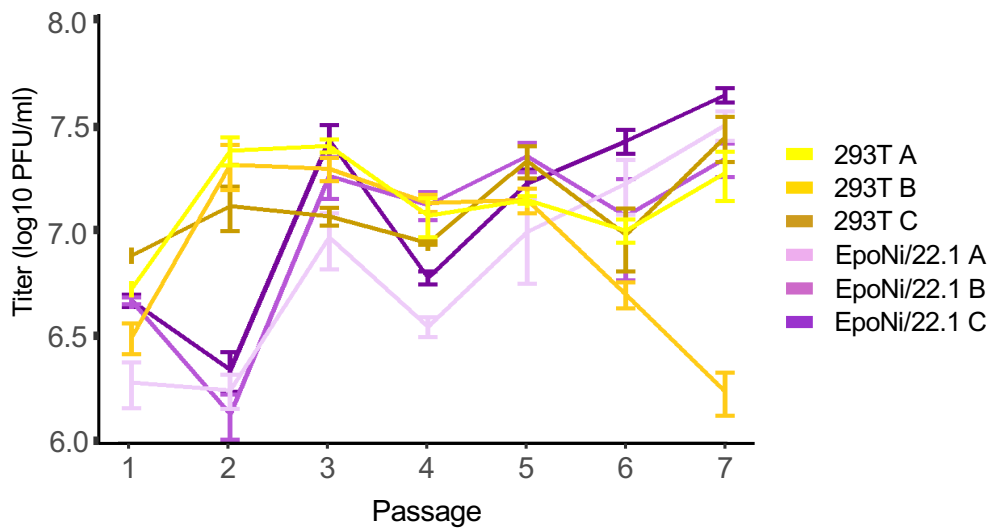
Data for virology assays were compiled and analyzed in Prism.

Supplemental Information

Species-Specific Evolution of Ebola Virus

**during Replication in Human and Bat Cells**

Zachary J. Whitfield, Abhishek N. Prasad, Adam J. Ronk, Ivan V. Kuzmin, Philipp A. Ilinykh, Raul Andino, and Alexander Bukreyev

Suppl. Fig. 1. Titer of recovered viral populations for each clone during the course of passaging.
Relevant to Fig. 1. The titer of clone 293T B drops dramatically during passages 5 and 6. Limit of detection 10 PFU/mL.
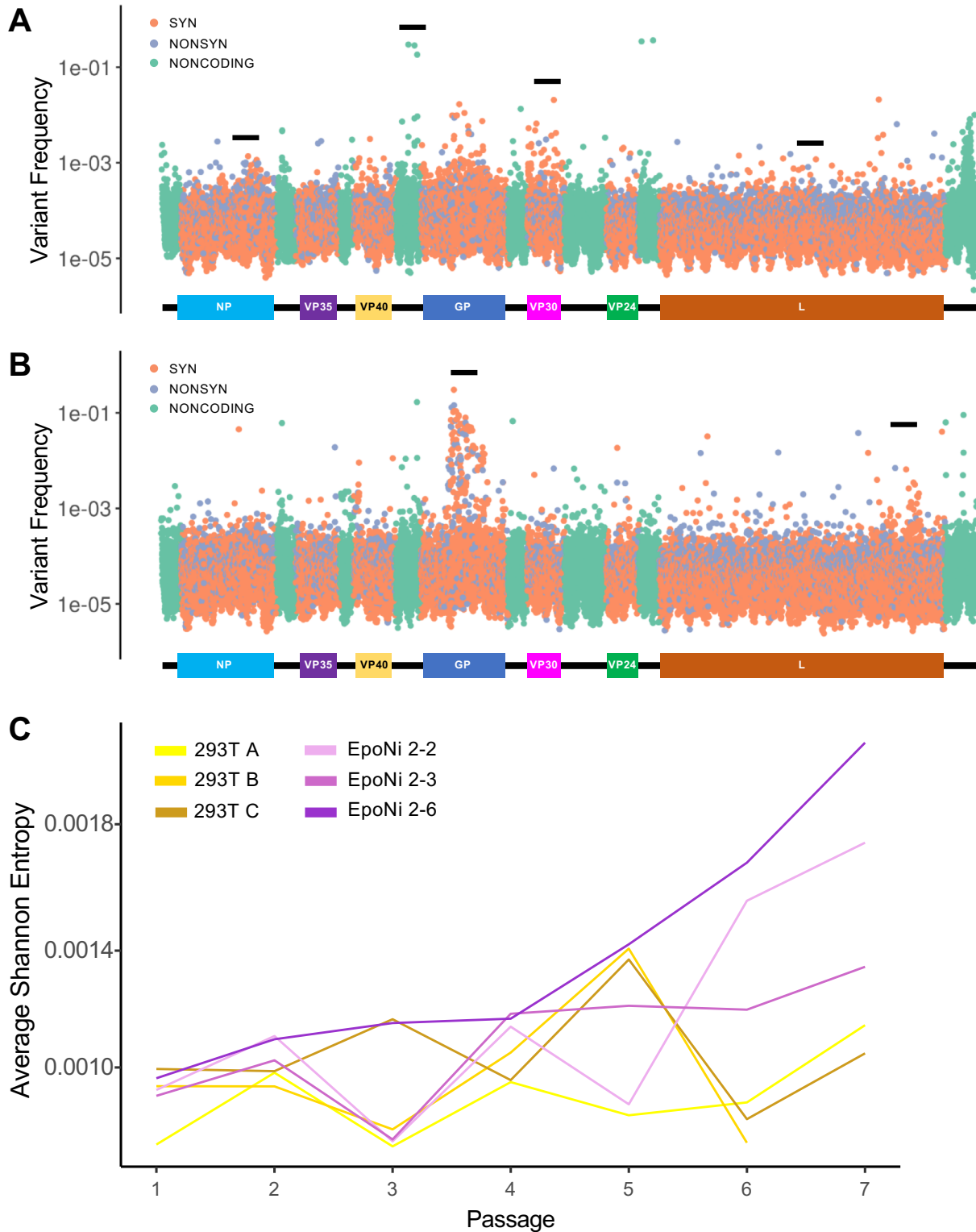
Suppl. Fig. 2. Additional characterization of passaged virus Relevant to Figs. 2 and 3. The average frequency for every variant in the EBOV genome across (A) 293T-derived and (B) EpoNi/22.1-derived virus populations at passage 7. CtoU(-/genomic)/GtoA(+/coding) mutations are not shown due to their high frequency to improve visual clarity. These mutations are extremely common, and would obscure the general pattern observed were they included. As some are not immediately apparent in these representative figures, black horizontal bars denote locations of variant clusters discussed in text. (C) The average Shannon Entropy increases faster in the EpoNi-derived viral genomes than in 293T-derived genomes. The increase in the average genome-wide Shannon entropy in the EpoNi/22.1-passaged lines was largely due to the region of GP's glycan cap and mucin-like domain, where a small, but measurable effect size could be detected (Cohen's d = 0.31 at passage 6).
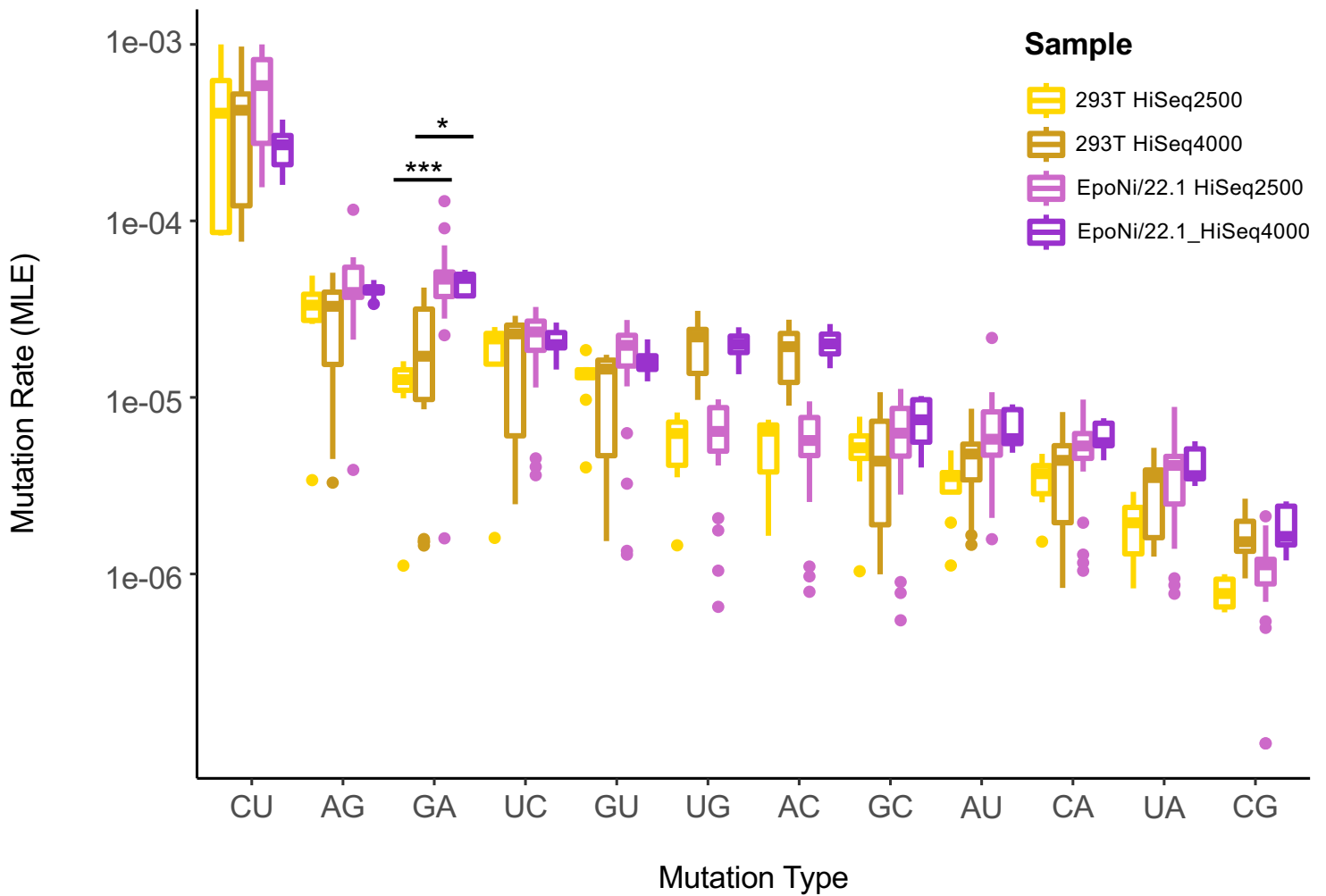
**Suppl. Fig. 3. Comparison of mutation frequency by host and sequencer. Relevant to STAR Methods.** Mutation frequencies for each type of nucleotide substitution, grouped by host cell-type and sequencer. Individual boxplots are made up of datasets corresponding to passages comprising data generated from individual runs on an Illumina HiSeq 2500 or HiSeq 4000. *: corrected p-value < 0.05 (GtoA in 293T vs. EpoNi/22.1 from HiSeq4000); ***: corrected p-value < 10^{-4} (GtoA in 293T vs. EpoNi/22.1 from HiSeq2500).
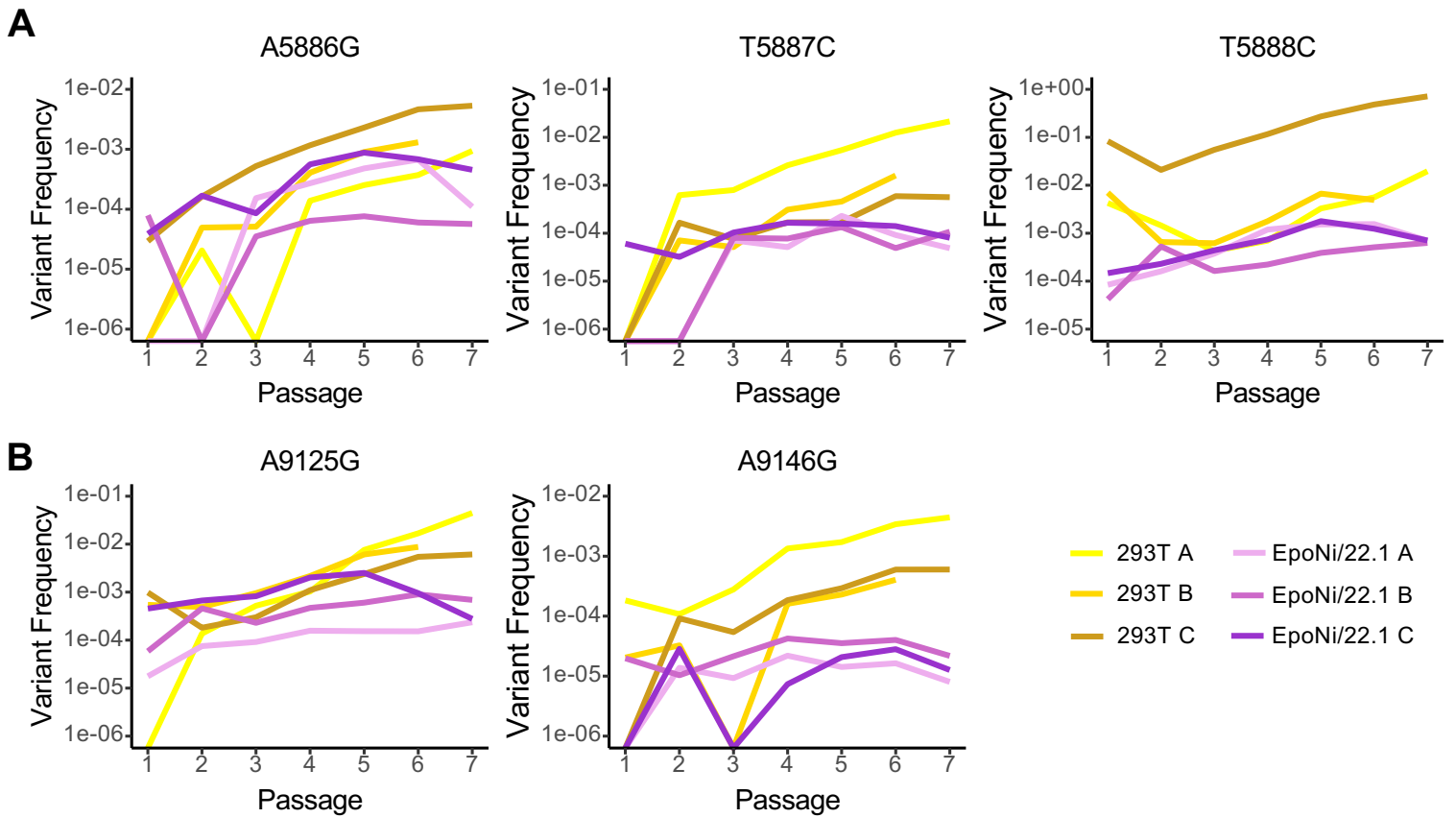
Suppl. Fig. 4. Trajectory of select high-fitness mutations in 293T-derived virus. Relevant to Figs. 2 and 5. Plots of frequency versus passage for mutations in the VP40 gene end (A) and VP30 (B). 293T replicates are in shades of yellow; EpoNi/22.1 replicates are in shades of purple.
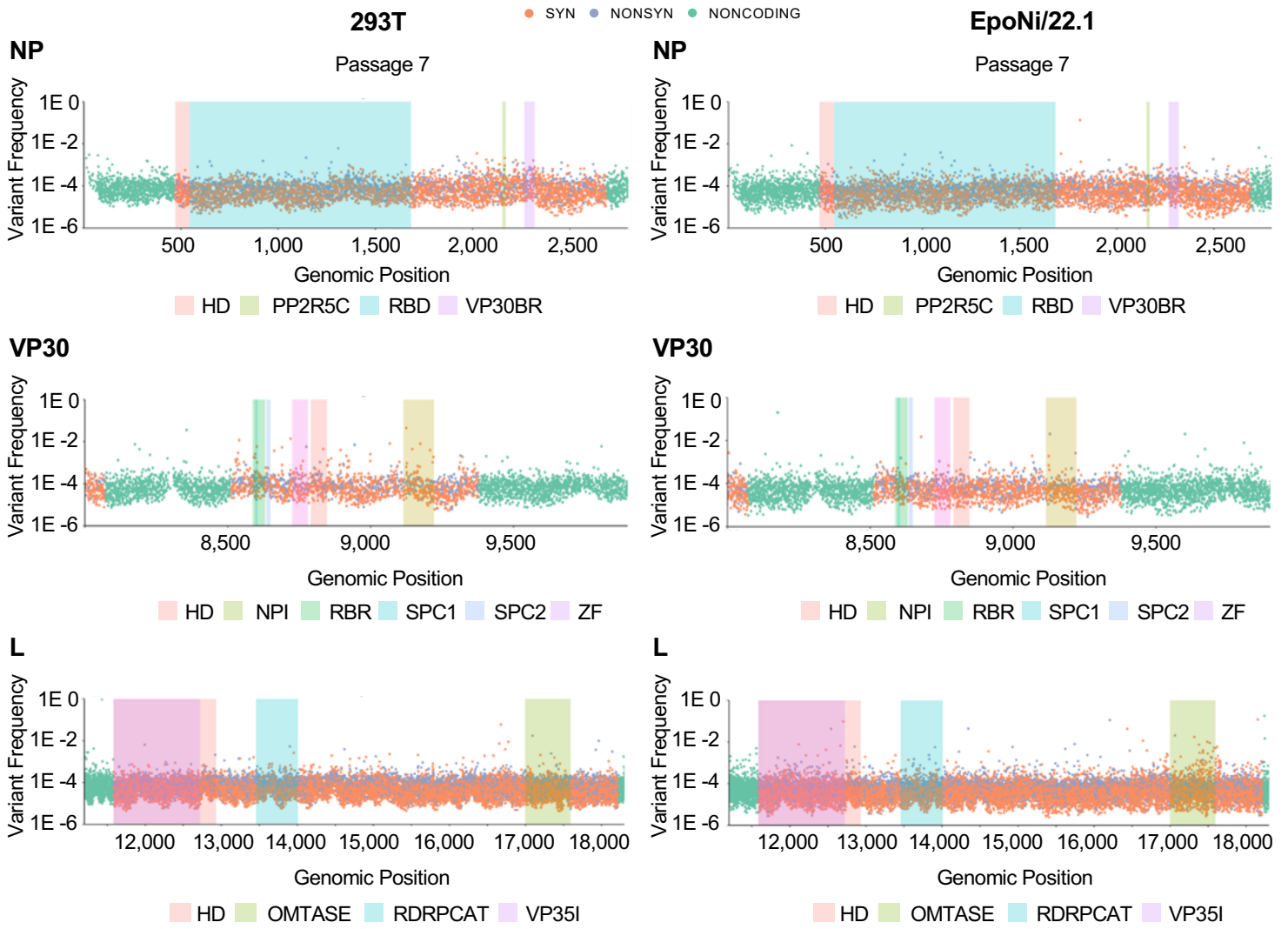
Suppl. Fig. 5. Clusters of variants detected in specific protein domains. Relevant to Figs. 2, 5, 6, and 7. Maximum variant frequency at passage 7 among all 293T or EpoNi/22.1 clones is plotted for NP, VP30, and L. Specific domains within each protein are indicated by shaded boxes and indicate the following: NP (homooligomerization domain; PPP2R5C binding motif; RNA binding domain; VP30 binding region), VP30 (homooligomerization domain; NP interaction; RNA binding region; serine phosphorylation cluster 1; serine phosphorylation cluster 2; zinc finger) and L (homooligomerization domain; SAM-dependent 2'-O-Mtase; RdRp catalytic domain; VP35 interaction).
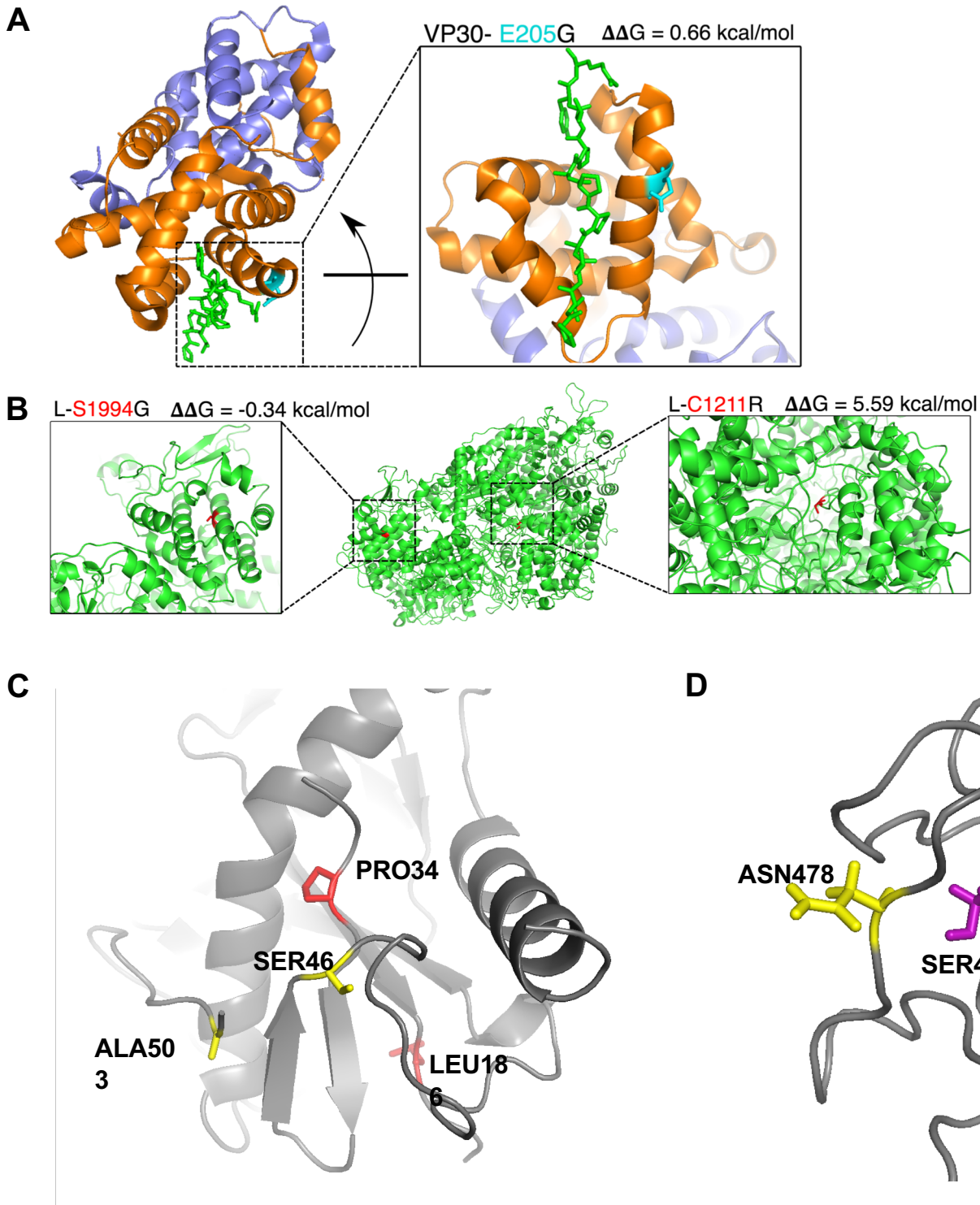
Suppl. Fig. 6. Stability predictions and mapping of newly identified and previously known mutations. Relevant to Figs. 2, 5, 6, and 7. (A) VP30-E205G (highlighted blue residue) on PDB structure 5T3T, visualized in PyMol. The stick structure is a portion of NP, which interacts with the alpha-helix to which E205 belongs. (B) The positions of L-S1994G and L-C1211R (highlighted in red) on the predicted structure of EBOV L protein (based on the VSV L protein). The predicted impact of each mutation, as determined by FoldX, is shown in kcal/mol. (C) Location of nonsynonymous mutations exhibiting positive fitness in 293T-passaged virus (red) which localized near mutations identified in the 2014 outbreak (yellow) in GP. (D) The same for NP. Purple denotes a positive fitness mutation (in at least one replicate) also identified in the 2014 outbreak.
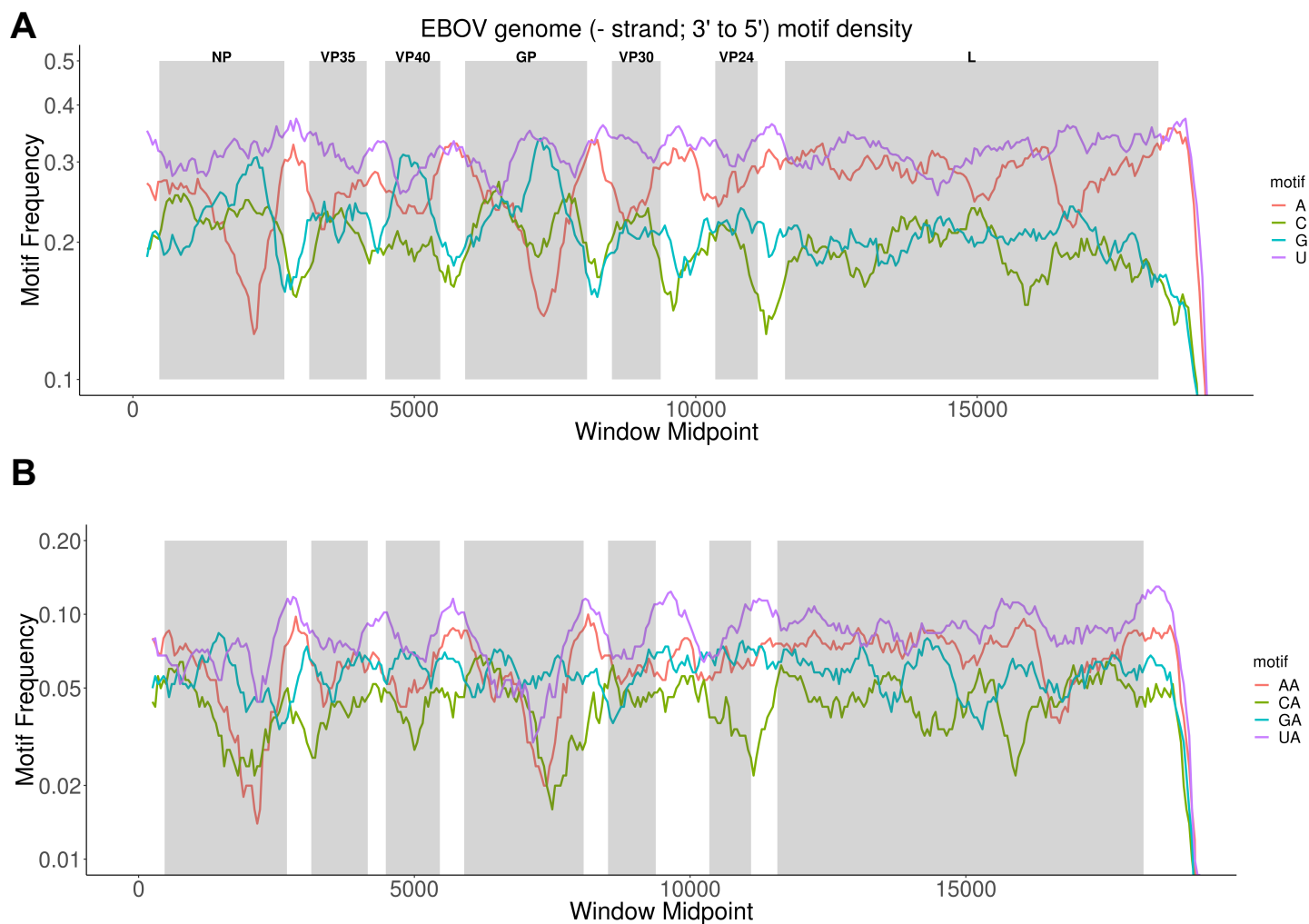
Suppl. Fig. 7. Nucleotide composition of the EBOV genome. Relevant to Figs. 3 and 4. Frequencies of individual nucleotides (A) or the dinucleotides 3'-XA-5' (B) are shown. The genome is displayed 3' to 5' with respect to the negative sense genomic strand. Motif frequencies were calculated with a window size of 500 nucleotides and a step size of 50 nucleotides.

Suppl. Table 1. ADAR qPCR primers (Relevant to Methods)

	Forward Primer	Reverse Primer
P. vampyrus ADAR1	ACTTTGAAAACGGCCAGTGG	TAGAAGGACGGCATCTCCATG
Human ADAR1	ATCAGCGGGCTGTTAGAATATG	AAACTCTCGGCCATTGATGAC
P. vampyrus 18s	CACGGCGACTACCATCGAA	CGGCGACGACCCATTC
Human 18s	GTAACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG
Standards		
P. vampyrus ADAR1	ACTTTGAAAACGGCCAGTGGGCCACCGACGACATCCCGGACGACCTGAACAGCATCCGCGCGGCCCCAGGCGAGTTCCGGGCCATCATGGAGATGCCGTCCTTCTA	
Human ADAR1	ATCAGCGGGCTGTTAGAATATGCCCAGTTCGCTAGTCAAACCTGTGAGTTCAACATGATAGAGCAGAGTGGACCACCCCATGAACCTCGATTTAAATTCCAGGTTGTCATCAATGGCCGAGAGTTT	
P. vampyrus 18s	CGGCGACGACCCATTCGAACGTCTGCCCTATCAACTTTCGATGGTAGTCGCCGTG	
Human 18s	GTAACCCGTTGAACCCCATTCGTGATGGGGATCGGGGATTGCAATTATTCCCCATGAACGAGGAATTCCCAGTAAGTGCGGGTCATAAGCTTGCGTTGATTAAGTCCCTGCCCTTTGTACACACCGCCCGTCGCTACTACCGATTGGATGG	

Suppl. Table 2. Competition assay PCR and sequencing primers (Relevant to Methods)

	Forward Primer	Reverse Primer
NP N5566S	CAGGCTTATTGATTGTCAAA	TGTCACTGTCCTGGTTCCTG*
GP L256P	AAGGTGTCGTTGCATTTCTG*	CTCGTGTTGGTGTTCTCTGC
VP30 E205G	AGTACCGTCAATCAAGGAGC	ATCAGACCATGAGCATGTCC*
L C1211R	CATCAACTCCTGTTATGAGT	GATCGTTGTACCTGTGAACA*
L S1992G	AGGTGCTGGTGCCTTACTAT	CGAATCTCTGCTCTAAGATG*

Asterisk indicates primer used for sequencing.