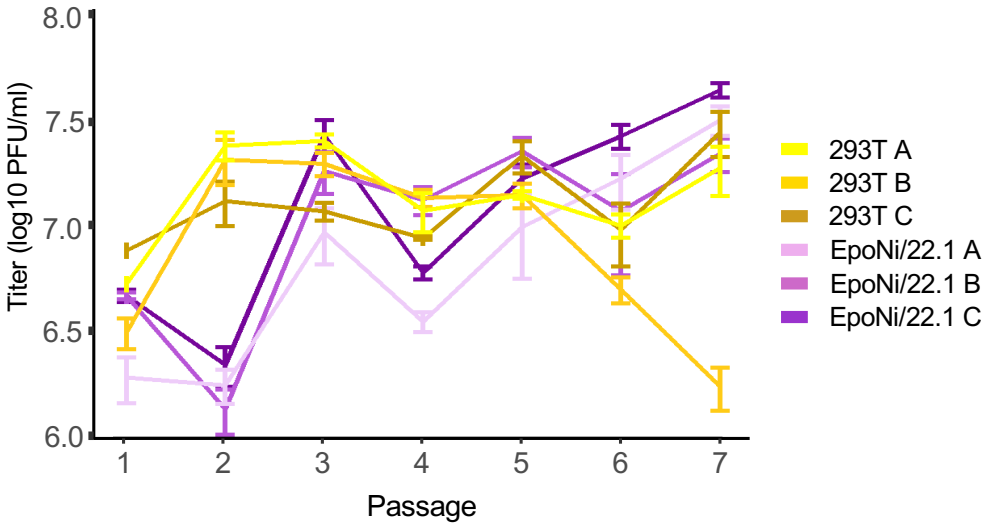
Cell Reports, Volume *32*

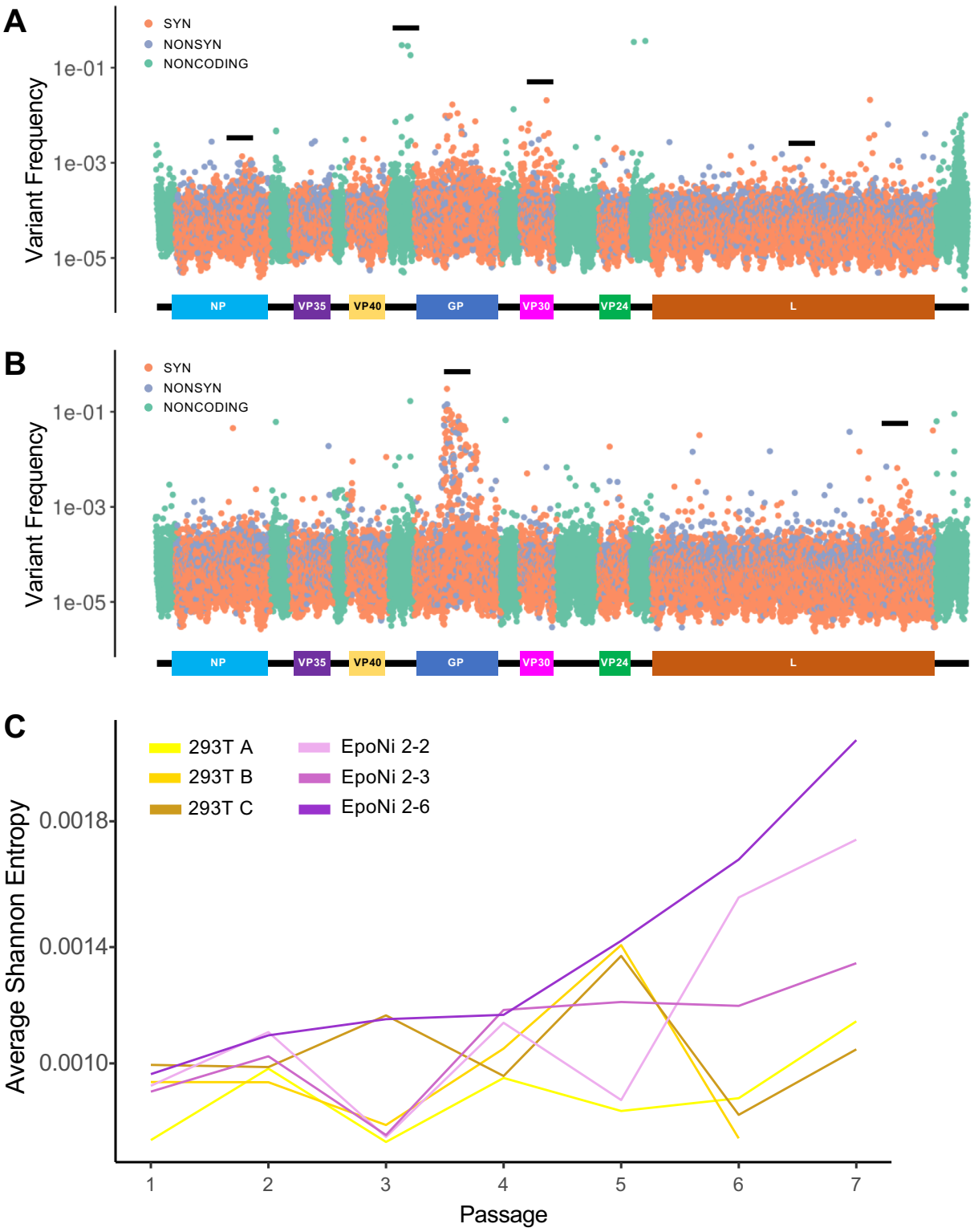
Supplemental Information

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

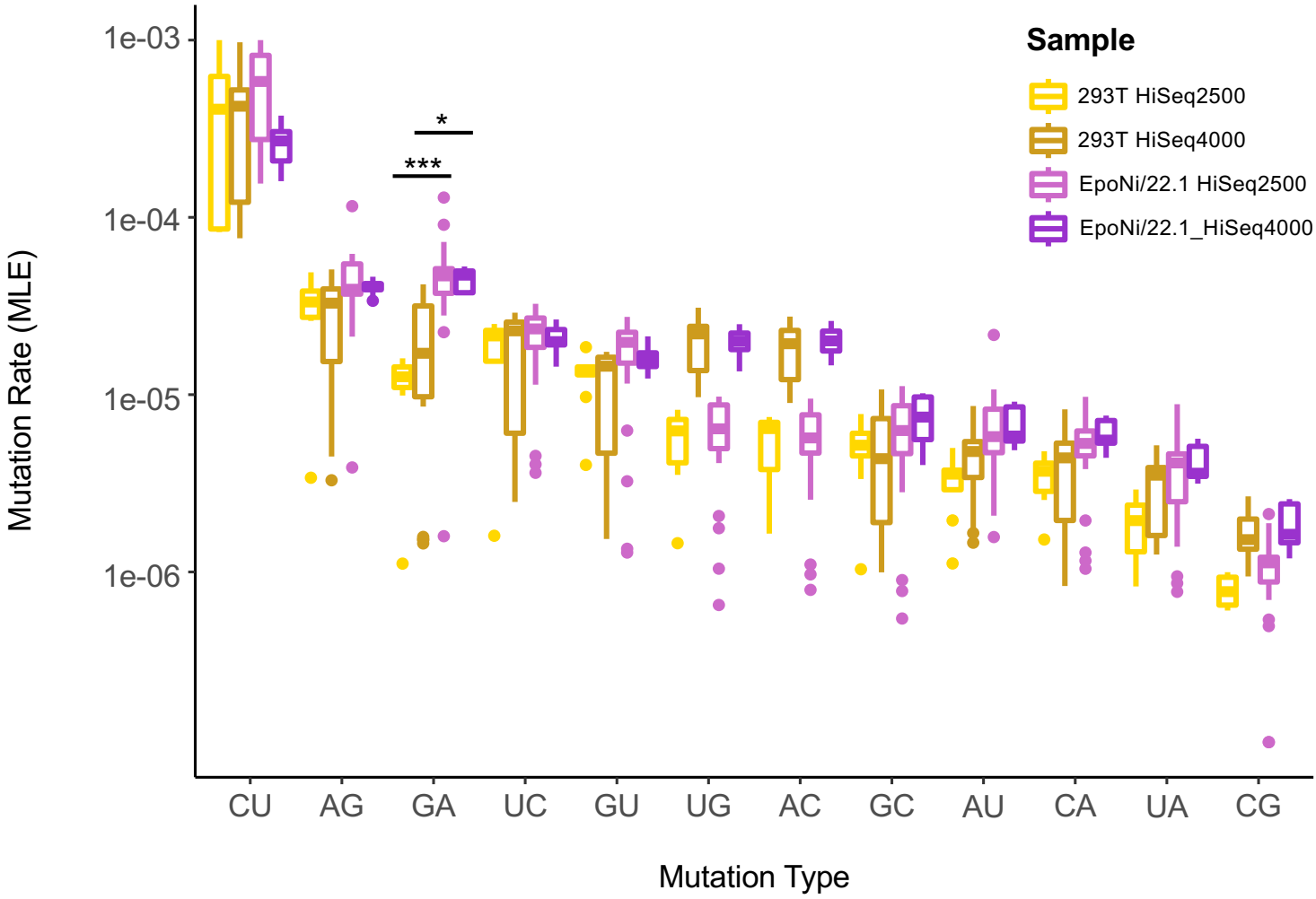
Zachary J. Whitfield, Abhishek N. Prasad, Adam J. Ronk, Ivan V. Kuzmin, Philipp A. Ilinykh, Raul Andino, and Alexander Bukreyev



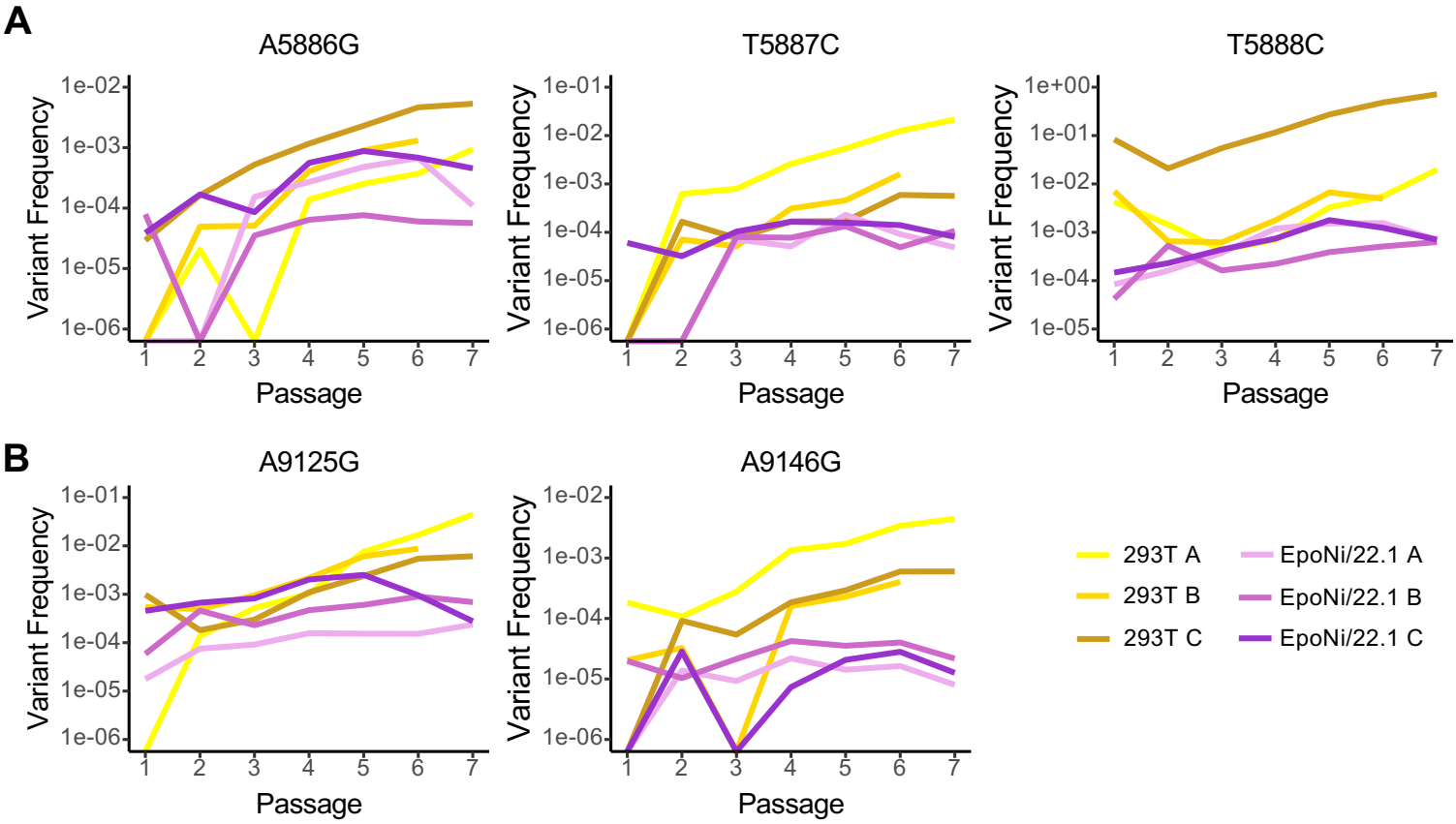
Suppl. Fig. 1. Titer of recovered viral populations for each clone during the course of passaging. Relevant to Fig. 1. The titer of clone 293T B drops dramatically during passages 5 and 6. Limit of detection 10 PFU/mL.



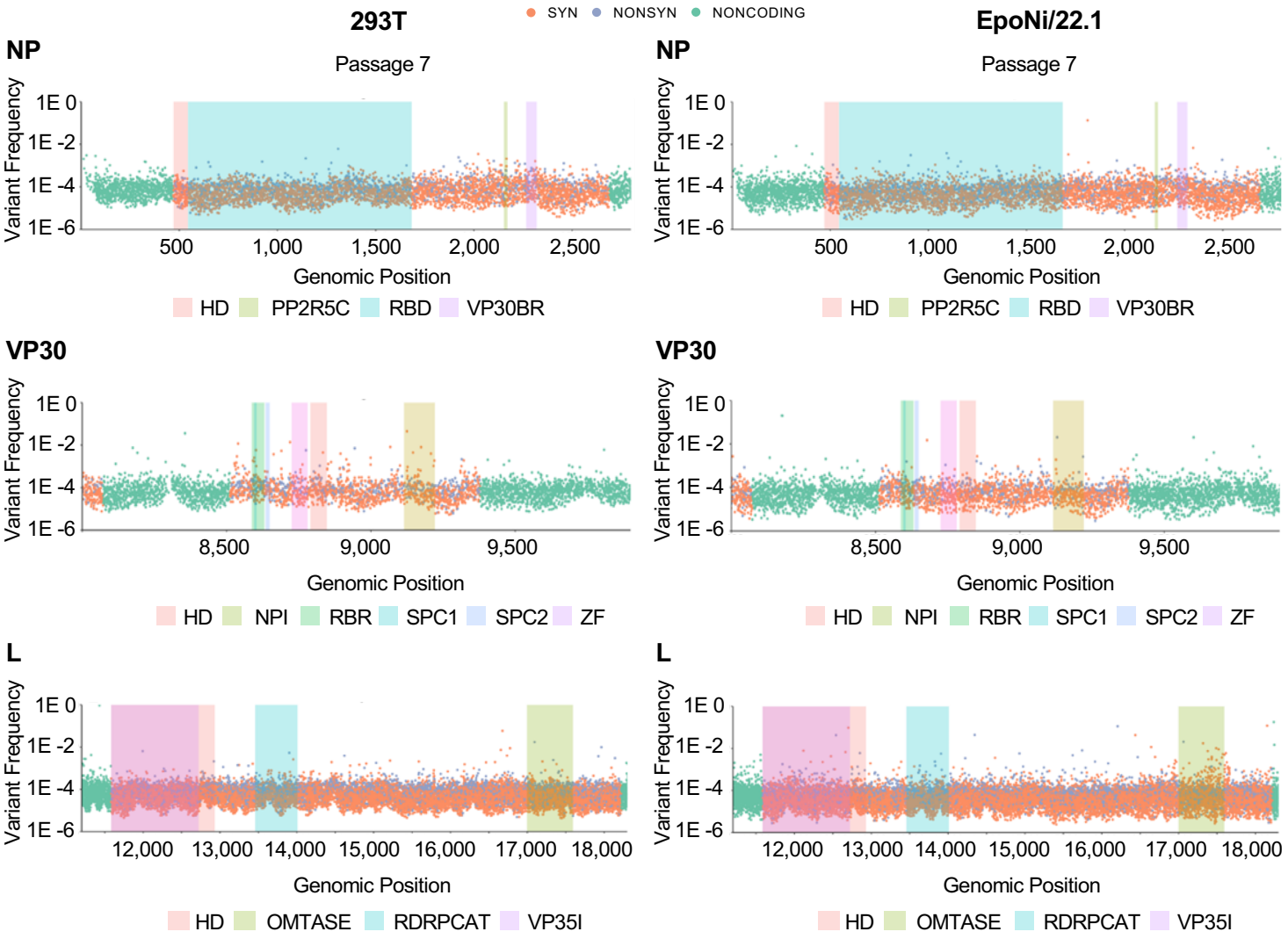
Suppl. Fig. 2. Additional characterization of passaged virus Relevant to Figs. 2 and 3. The average frequency for every variant in the EBOV genome across (A) 293T-derived and (B) EpoNi/22.1-derived virus populations at passage 7. CtoU(-/genomic)/GtoA(+/coding) mutations are not shown due to their high 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).



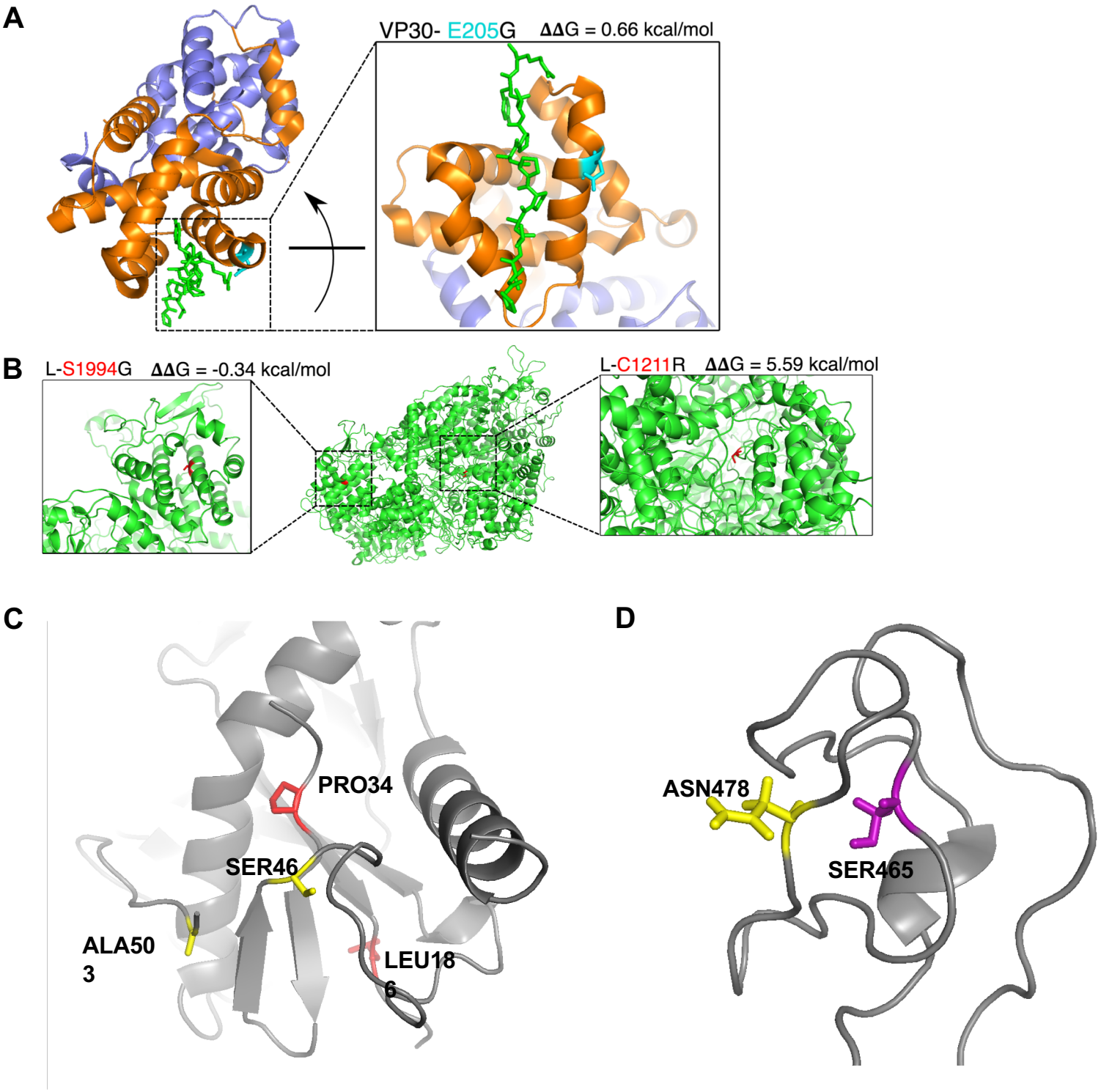
Suppl. Fig. 3. Comparison of mutation frequency by host and sequencer. Relevant to STAR Methods. Mutation frequencies for each type of nucleotide substitution, grouped by host cell-type and sequencer. Individual boxplots are made up of datasets corresponding to passages comprising data generated from 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).



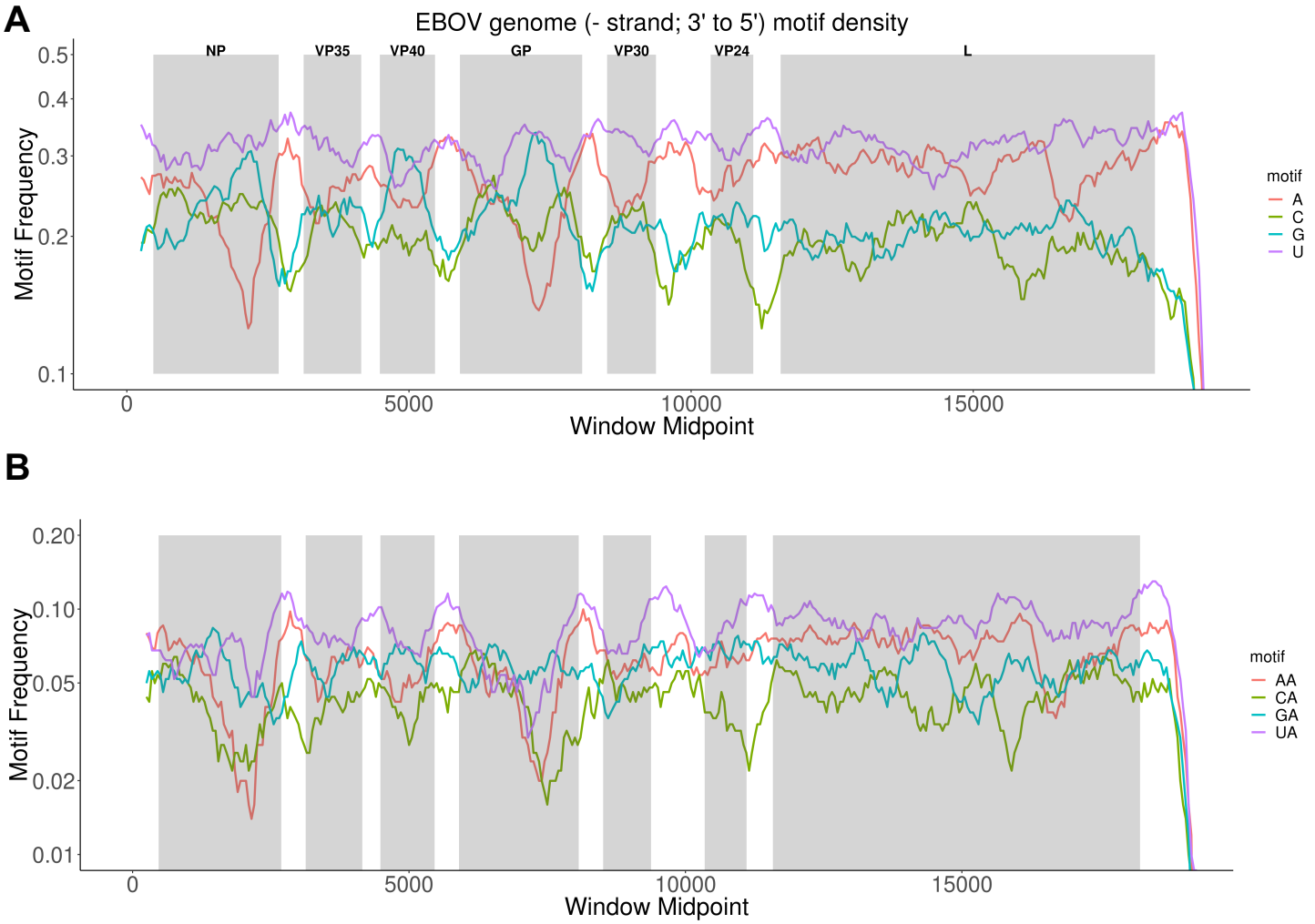
Suppl. Fig. 4. Trajectory of select high-fitness mutations in 293T-derived virus. Relevant to Figs. 2 and 5. Plots of frequency versus passage for mutations in the VP40 gene end (A) and VP30 (B). 293T replicates are in shades of yellow; EpoNi/22.1 replicates are in shades of purple.



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.