Genes to Cells

Functional mutations in spike glycoprotein of Zaire ebolavirus associated with an increase in infection efficiency

Mahoko Takahashi Ueda[1,*†], Yohei Kurosaki[2,†], Taisuke Izumi[3,4], Yusuke Nakano[3], Olamide K. Oloniniyi[2,5], Jiro Yasuda[2,5], Yoshio Koyanagi[3], Kei Sato[3,4*] and So Nakagawa[1,6*]

[1]*Micro/Nano Technology Center, Tokai University, 411 Kitakaname, Hiratsuka, Kanagawa 259-1193, Japan*
[2]*Department of Emerging Infectious Diseases, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan*
[3]*Laboratory of Systems Virology, Institute for Frontier Life and Medical Sciences, Kyoto University, 53 Shogoinkawahara-cho, Sakyo-ku, Kyoto, 606-8507, Japan*
[4]*CREST, Japan Science and Technology Agency, Saitama 322-0012, Japan*
[5]*Graduate School of Biomedical Sciences and Program for Nurturing Global Leaders in Tropical and Emerging Communicable Diseases, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan*
[6]*Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan*

Ebola virus (EBOV) is extremely virulent, and its glycoprotein is necessary for viral entry. EBOV may adapt to its new host humans during outbreaks by acquiring mutations especially in glycoprotein, which allows EBOV to spread more efficiently. To identify these evolutionary selected mutations and examine their effects on viral infectivity, we used experimental–phylogenetic–structural interdisciplinary approaches. In evolutionary analysis of all available *Zaire ebolavirus* glycoprotein sequences, we detected two codon sites under positive selection, which are located near/within the region critical for the host-viral membrane fusion, namely alanine-to-valine and threonine-to-isoleucine mutations at 82 (A82V) and 544 (T544I), respectively. The fine-scale transmission dynamics of EBOV Makona variants that caused the 2014–2015 outbreak showed that A82V mutant was fixed in the population, whereas T544I was not. Furthermore, pseudotype assays for the Makona glycoprotein showed that the A82V mutation caused a small increase in viral infectivity compared with the T544I mutation. These findings suggest that mutation fixation in EBOV glycoprotein may be associated with their increased infectivity levels; the mutant with a moderate increase in infectivity will fix. Our findings showed that a driving force for Ebola virus evolution via glycoprotein may be a balance between costs and benefits of its virulence.

Introduction

Ebola virus (EBOV) is a member of the family *Filoviridae*, which are nonsegmented, negative-sense RNA viruses. This virus is transmitted by direct contact with the blood or body fluids of infected individuals, and its infection often causes severe hemorrhagic fevers with multi-organ failure in various primates, including humans (Feldmann & Geisbert 2011). The viral particle consists of seven structural proteins: nucleoprotein (NP), RNA-dependent RNA polymerase cofactor (VP35), viral matrix protein 40 (VP40), glycoprotein (GP), transcriptional activator 30 (VP30), the minor matrix protein 24 (VP24) and RNA-dependent RNA polymerase (L) (Feldmann

Communicated by: Haruhiko Siomi
**Correspondence:* mahoko@tokai.ac.jp, ksato@virus.kyoto-u.ac.jp, and so@tokai.ac.jp
†These authors equally contributed to this study.

DOI: 10.1111/gtc.12463

et al. 1993). GP is the only protein found on the surface of the EBOV virion, and thus, can act as a major antigen that is recognized by host immunity and can be a target for vaccines and therapeutic agents.

Glycoprotein, which mediates viral entry into the host cell, is 676 amino acids in length and is post-translationally cleaved at the amino acid position 501 by the cellular protease furin to yield the GP1 and GP2 subunits (Fig. 1, top panel; Volchkov *et al.* 1998). GP1 is a surface subunit that is responsible for cell attachment, whereas GP2 is a transmembrane subunit that catalyzes membrane fusion (Fig. 1, top). The trimeric structure of EBOV GP has the shape of a chalice, with GP1 and GP2 forming the cup and base, respectively (see Fig. 2A) (Lee *et al.* 2008). After entering into the host cell, GP is further processed and the mucin-like domain (MLD) and glycan cap in GP1 are cleaved by cathepsins in the low-pH environments of the late endosome (Fig. 2A, dark gray regions) (Chandran *et al.* 2005; Schornberg *et al.* 2006). This primed GP form (Fig. 2B) binds to the Niemann-Pick C1 (NPC1) membrane protein, which is an essential receptor for EBOV entry (Miller *et al.* 2012). A series of conformational changes of the primed GP occur after its binding to NPC1 (Bale *et al.* 2012; Brecher *et al.* 2012), resulting in the exposure of an internal fusion loop (IFL) that is normally buried in GP1 of preprimed GP (Lee *et al.* 2008). Subsequent insertion of IFL to the host membrane triggers membrane fusion.

In central Africa, *Zaire ebolaviruses* (ZEBOV) has caused multiple outbreaks of Ebola virus diseases as the virus was first discovered in the Democratic Republic of the Congo in 1976 (EBOV variant Mayinga). During a very recent outbreak in West Africa that started in December 2013 in Guinea, which was the largest ever recorded, the number of infected cases was more than 28 000. This outbreak was caused by a different variant of EBOV, known as the Makona variant (Kuhn *et al.* 2014). Epidemiological study and genetic analysis showed that the Makona variants spread by human-to-human transmission following a single virus introduction from an unidentified natural reservoir in a forest area of

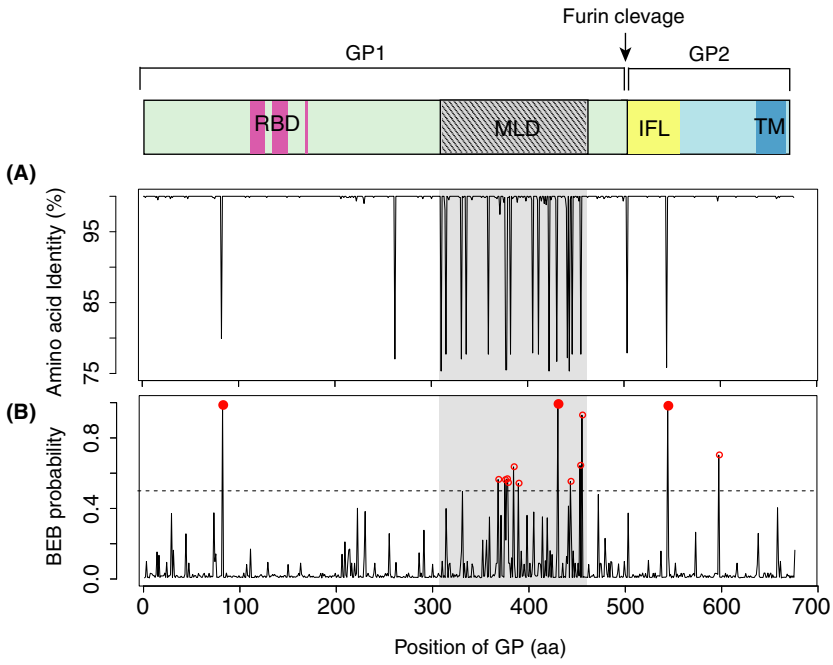

Figure 1 Amino acid variations in the GP domain structures of EBOV strains. Upper panel: schematic of EBOV GP structure; RBD, receptor-binding domain (magenta); MLD, mucin-like domain (hatched gray); IFL, internal fusion loop (yellow); TM, transmembrane domain (blue). Domain positions are based on the crystal structure 3CSY (Lee *et al.* 2008). Lower panel A: Percentage identities of amino acids in GP domain structures. For calculation of identity, prototype strain Mayinga sampled in 1976 (NC_002549) was used as reference sequence. Lower panel B: GP amino acid sites with BEB posterior probabilities for positive selection under the M8 model of PAML. Amino acid sites with a BEB posterior probability of >0.5 (open red circle) and >0.95 (filled red circle) are shown. Dashed line represents a BEB of 0.5.

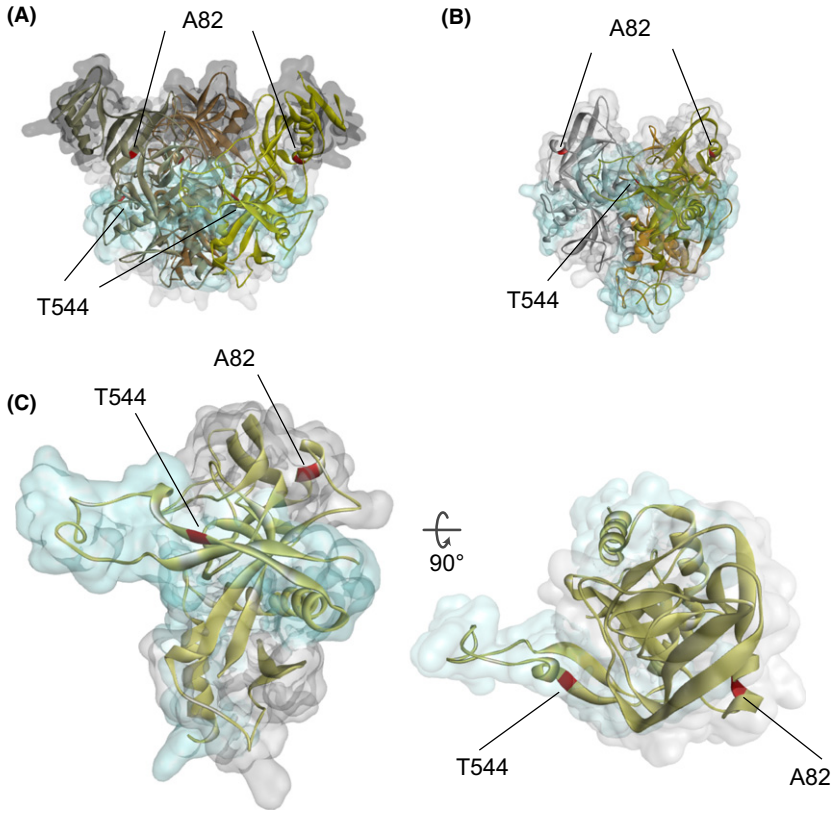

Figure 2 Protein structure models of EBOV GP with A82V and T544I mutations. Molecular surface with ribbon representations for A: Trimeric GP (3CSY) and B: primed GP (5HJ3). GP1 glycan cap, which is removed by endosomal proteases (cathepsins), and GP2 subunit are shown in dark gray and light blue, respectively. C: Top and lateral views of host-primed GP monomer (5F1B). The residues at position 82 and 544 are shown in red.

Guinea (Baize *et al.* 2014). High-depth sequencing data from patient blood samples are available for the ZEBOV outbreaks, in particular for the most recent outbreak (Gire *et al.* 2014; Park *et al.* 2015). These data offer an unprecedented opportunity to identify genomic regions under selection in the ZEBOV genome. Previous studies using the ZEBOV genomic sequences showed that GP is highly variable compared with other genes (Azarian *et al.* 2015; Liu *et al.* 2015; Park *et al.* 2015). Considering the function of the ZEBOV GP during cell entry and host membrane fusion, mutations in this protein may affect its ability to bind interaction partners, such as cell receptors. ZEBOV sequence data are being accumulated day to day, and currently, more than a thousand of the sequences are available. In this study, to identify functional mutations in GP, we conducted a collaborative investigation of evolutionary, structural and experimental–virology analyses using all available sequences (as of January 19, 2016). In our study, we successfully identified two functional mutations that elevate infection efficiency. We believe that our approach will provide insights into the evolutionary dynamics of virus–host interactions of Ebola virus.

Results

Molecular evolutionary analysis of ZEBOV

To investigate the evolutionary dynamics of ZEBOV, we obtained 1232 ZEBOV genome sequences from NCBI GenBank. Out of these genomes, we extracted GP gene sequences to generate a core dataset containing 147 representative nucleotide sequences (Fig. S1 in Supporting information). We aligned the amino acid sequences with MAFFT L-INS-i (Katoh & Standley 2013) and plotted the percent identity (PI) to obtain an overview of the sequence diversity (Fig. 1A). GP, in particular GP2, showed high conservation with a PI of 98.8%. The average PI within

MLD (~90.3%) was lower than that outside MLD (~99.0%) as previously reported (Sanchez *et al.* 1998; Wertheim & Worobey 2009; Azarian *et al.* 2015; Liu *et al.* 2015; Park *et al.* 2015). However, in this study, we focused on residues outside the MLD because it is not sufficiently ordered to be observed in the crystal structure.

Next, we investigated whether positive selection acts on GPs using the site model of the PAML program, version 4.8 (Yang 2007). Likelihood scores were compared between a null model (M1a and M7) and a selection model (M2a and M8) using the likelihood ratio test (LRT). Both the selection models M2a and M8 were significantly favored ($P < 0.001$; LRT), we examined the posterior probability of Bayes empirical Bayes (BEB) method implemented in PAML (Tables S1 and S2 in Supporting information) and determined sites with BEB probability of >0.95 to be under positive selection. Three amino acid sites at the positions 82, 430 and 544 were identified by both comparisons, and the results were partially consistent with previous studies. In particular, the two sites at 82 (Ladner *et al.* 2015) and 430 (Azarian *et al.* 2015; Liu *et al.* 2015) were detected to have been under positive selection using several different methods, whereas the site at 544 was first identified in this study (see Table S3 in Supporting information for comparison of results with different methods). We then focused on the two mutations at positions 82 and 544 that are located outside MLD for further analysis (Fig. 1A and B).

Consistent with previous studies (Azarian *et al.* 2015), phylogenetic analysis using ZEBOV GP sequences showed that each epidemic was divided into groups (Fig. S1 in Supporting information). The mutation of alanine (A) to valine (V) at position 82 was found only in the Makona strain of the 2014–2015 outbreak, as described in a previous study (Ladner *et al.* 2015). In contrast, the mutation of threonine (T) to isoleucine (I) at 544 has frequently occurred during the past outbreaks (Fig. S2 in Supporting information). All filoviruses other than ZEBOV also had A and T at the sites corresponding to the ZEBOV position 82 and 544, respectively (but not Reston for the site corresponding 544). This indicates that these residues might start from the ancestral state 'A' and 'T' at the outset of each outbreak, respectively (Fig. S2 in Supporting information). To understand the detailed ZEBOV transmission history, we further investigated phylogenetic trees using the ZEBOV whole genome from the 2014 to 2015 outbreak (Figs 3, S3 and S4 in Supporting information). We first extracted 805 ZEBOV genomes without 'N' (undetermined nucleotide) of 1,232 genomes from the recent outbreak. We then constructed the phylogenetic tree based on the genome sequences using RAxML (Stamatakis 2006). The whole-genome phylogenetic analysis enables a finer-scale investigation of ZEBOV transmission history than single-gene analysis. The A82V mutation emerged around May 2014—about six months after the first case was reported—during the transmission between Guinea and Sierra Leone, which is consistent with a previous study (Ladner *et al.* 2015) (Fig. 3). After the occurrence of the mutation, the residue V was spread and fixed in the population. Very early strains in the recent outbreak contain a codon 'GCG' for alanine at position 82, whereas all ZEBOV strains in the past outbreaks contain a codon 'GCA' for the same amino acid, indicating the possibility that the nucleotide for the third position of the codon for A82 had already mutated in natural hosts before ZEBOV appeared in the human population during 2013.

However, the T544I mutation displays a different transmission history from the mutation at 82. The T544I mutation had also emerged during the early stage of the outbreak in Guinea, but was not fixed in the population (Fig. 3). We found only seven Makona strains with the T544I mutation, and six of those were collected from the patient blood samples during March 2014 in Guinea. In these mutants collected in Guinea, the A82V and T544I mutations did not occur concurrently. However, one strain obtained in Sierra Leone (collection date was unknown) contained both mutations at positions 82 and 544 (Fig. 3).

Infectivity changes induced by the two mutations

The residue at position 82 is located on an alpha-helix domain of GP1 far from the surface of the pre-primed GP (Fig. 2A). However, the residue is located very close to the protein surface in the primed GP and is particularly close to the NPC1 receptor-binding domain (Fig. 2B). Indeed, quite recently, several studies of this mutation independently reported that the A82V mutation increases ZEBOV entry efficiency in human cells (Diehl *et al.* 2016; Dietzel *et al.* 2016; Urbanowicz *et al.* 2016). The residue at 544 is located on IFL, which plays a key role in membrane fusion (Gregory *et al.* 2011). Our structural analysis based on the two ZEBOV crystal structures, namely preprimed GP [PDB ID: 3CSY (Lee *et al.* 2008)] and

Genes to Cells published by Molecular Biology Society of Japan and John Wiley & Sons Australia, Ltd.

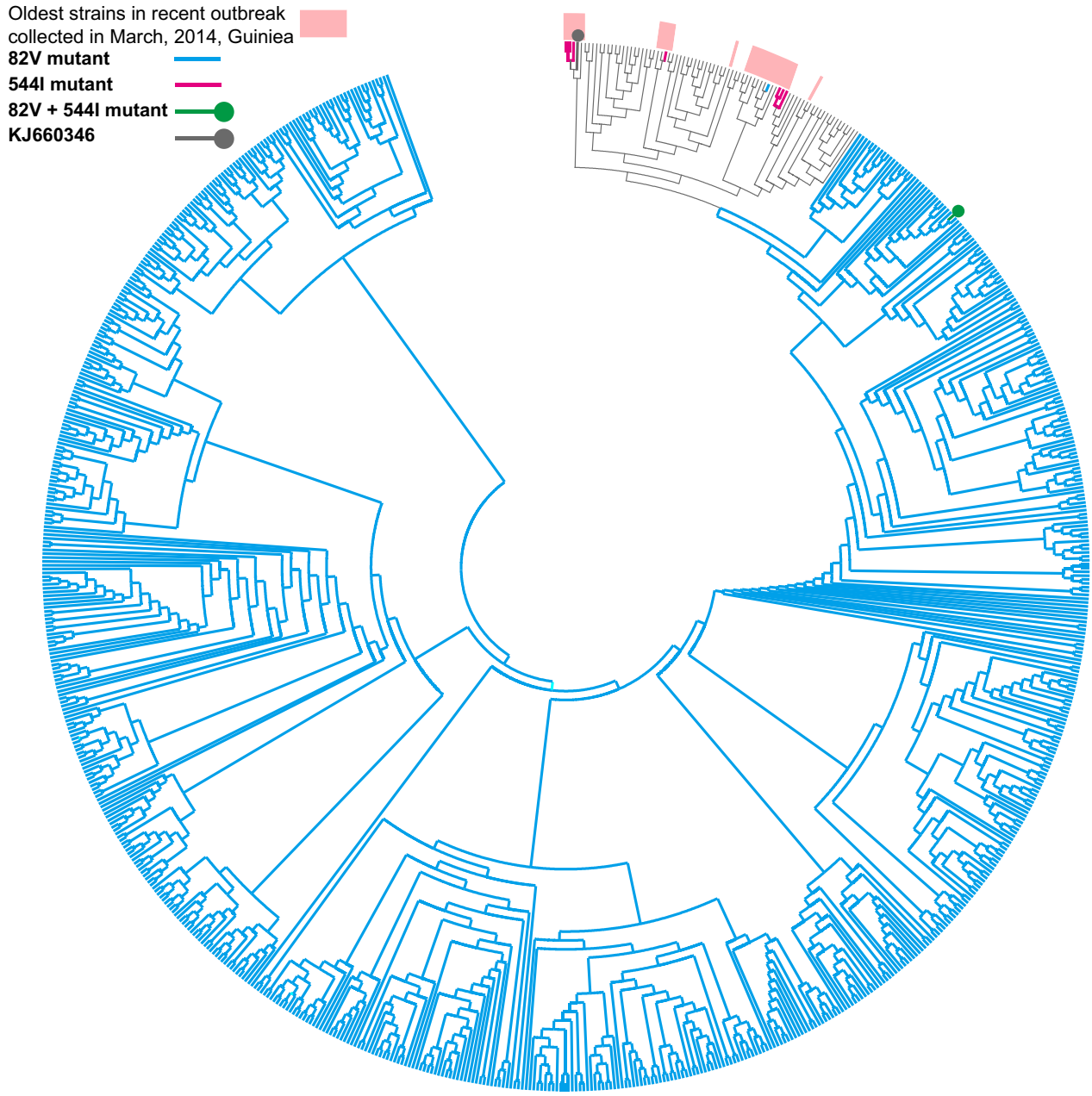

Figure 3 Whole-genome phylogenetic tree from 2014 to 2015 outbreak. ML phylogenetic trees using 805 ZEBOV genomes. Each node represents a genome of the ZEBOV strain. External nodes of very early strains collected in Guinea in May 2014 are colored red. Mutants are indicated by colored branches. A82V and T544I mutants are cyan and magenta, respectively. An ancestral strain used in the experiment and double mutant (A82V/T544I) is indicated by gray and green branches with circles, respectively.

postprimed GP [PDB ID: 5F1B (Wang *et al.* 2016)], showed changes in the protein stability induced by these mutations. The structure models with V and I at positions 82 and 544, respectively (A82V and T544I mutants), were generated, and the GP stabilities for these mutants were calculated by CHARMm force fields. We found that each mutation affects the protein stability of postprimed GP (mutation energy: 1.08 or −1.47 kcal/mol for A82V or T544I, respectively), but not the preprimed GP.

This indicates the possibility that these mutations can alter host–virus interaction in the late endosome.

To investigate whether these two mutations affect the host–virus interaction, we conducted virological experiments using cell cultures. We prepared pseudotyped vesicular stomatitis virus (VSV) particles with GP of the Makona strains and its derivatives, and measured their entry ability into human hepatoma Huh-7 cells. The GP sequence used in this study originated from a very early strain detected in Guinea during March 2014 (EBOV/H.sapiens-wt/GIN/2014/Makona-Gueckedou-C15, GenBank/ENA/DDBJ accession number: KJ660346, see Fig. 3), which contains A and T at positions 82 and 544, respectively (A82/T544). Based on this GP sequence, we constructed Makona GP mutants by substituting A with V at 82 (V82/T544), or T with I at 544 (A82/I544) or a concurrent mutant at both sites (V82/I544). Our GP sequence comparison among filoviruses showed that the ancestral amino acids of Makona GP at the 82 and 544 positions were A and T, respectively (Fig. S2 in Supporting information). Therefore, we assessed viral infectivity of each GP variant by comparing them to that of the ancestral GP (A82/T544). The mutation A to V at position 82 (V82/T544) increased viral infectivity 1.8-fold more than that of the ancestral Makona variant (Fig. 4, upper panel). The pseudotyped viruses with single mutation T to I at position 544 (A82/I544) infected the host cells more efficiently (4.3-fold) than those with the mutation at 82 (V82/T544). In addition, the concurrent mutant at both positions 82 and 544 (V82/I544) further enhanced infectivity by 8.6-fold compared with the ancestral Makona variant (Fig. 4, upper panel). This effect was not related to differences in the amount of pseudotyped viruses inoculated to the Huh-7 cells or GPs incorporated into virions, as VSV matrix protein M and GPs in the producer cells or the released virions were detected in similar amount in spite of GP mutations (Fig. 4, lower panel). We also observed the significant increase in Makona GP-mediated infection by T544I, and weak increase by A82V in three other human cell lines from different tissues (Fig. S5 in Supporting information). These results suggested that mutations A82V and T544I mutations in GP of Makona variants increased viral infectivity.

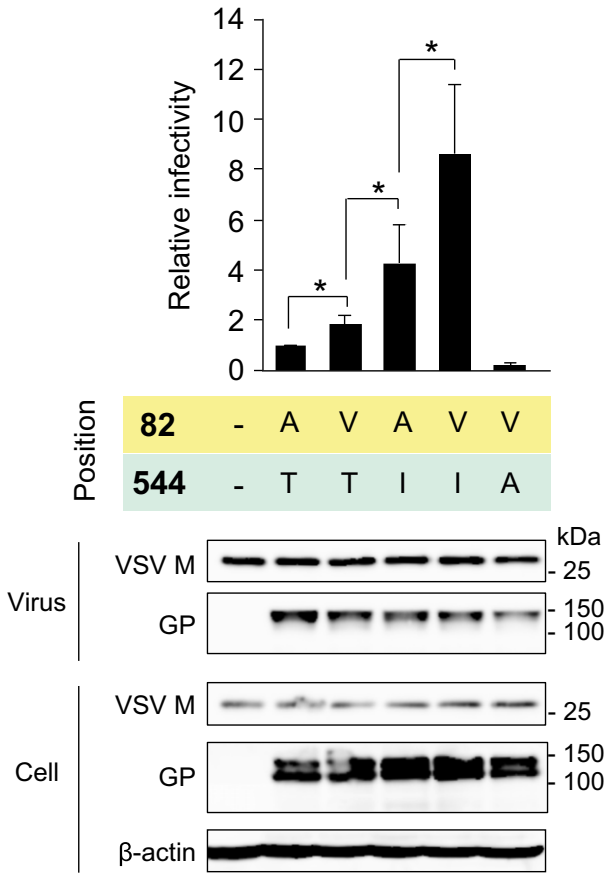

Figure 4 Infectivity of pseudotyped VSVs mediated by GP from the Makona variant or its A82V and T544I mutants. Equal volumes of pseudotyped VSVs carrying Makona GP with ancestral amino acid residues at position 82 and 544 (A82/T544) and the indicated mutants were used to infect to human hepatoma Huh-7 cells. 'Bald' virus particles prepared by transfecting empty vector, pCAGGS, to producer cells were used as noninfectious control. Transduced cells were quantified by luciferase assay at 24 h postinfection. Relative infectivity of pseudotype virus with each GP mutant was determined relative to the ancestral Makona GP A82/T544 (upper panel). Data represent means ± standard deviation of three independent experiments. Significance by Student's t-test: *, $P < 0.05$. Makona GP and VSV M incorporated into the virus particles or expressed in the lysates of the pseudotype virus producing 293T cells were detected by immunoblot using anti-GP or VSV M monoclonal antibodies (lower panel). Residues at positions 82 and 544 for upper and lower panels are shown in the middle in yellow and blue boxes, respectively.

Discussion

In the present study, we found two positively selected sites at positions 82 and 544 using maximum likelihood analysis and showed their influence on the virus infectivity to human cells. Our results showed that both mutations in the Makona strain increased ZEBOV infectivity in hepatoma Huh-7 cells

Genes to Cells published by Molecular Biology Society of Japan and John Wiley & Sons Australia, Ltd.

compared with that of the ancestral virus (Fig. 4). The mechanism of how these mutations increase the infectivity is unknown. For the A82V mutation, one possible explanation is alteration of GP-NPC1 binding. NPC1 is an essential host factor for filovirus entry and infection (Carette *et al.* 2011; Côté *et al.* 2011) and is known to be localized in late endosomes and lysosomes (Carstea *et al.* 1997; Cruz *et al.* 2000; Davies & Ioannou 2000). A hydrophobic cavity on the head of primed GP engages the C-terminal domain of NPC1. Specifically, residues near the position 82, such as 79 (V), 80 (P) and 83 (T), are part of the cavity and interact with the key residues in the NPC1 C-terminal domain (Wang *et al.* 2016) (Fig. 5A). The residue at 82 is located under the residue at 79, which contributes to the tight hydrophobic interactions with the residue 506 of NPC1 (Fig. 5A). Our analysis of the accessible surface area of primed GP showed that the A82V mutation increased the surface accessibility of residue 79 by 2.3%. Only the residues at 79, 80 and 83 near the position at 82 can interact with a residue at 506 of NPC1. Thus, if the distortion around the residue 82 occurs, the interaction between the three residues in primed GP and the residue at 506 in NPC1 may be altered and result in the increase in the infectivity (Fig. 5B and C).

The influence of the A82V mutation on EBOV infectivity was assessed using five cell lines from other mammalian species, and this mutation was found to increase the infectivity only in primate cells (Diehl *et al.* 2016; Urbanowicz *et al.* 2016). However, as described previously, A82V was a unique mutation during the recent outbreak, which was the largest and longest outbreak in history. This epidemic, lasting two years, might have provided sufficient time to generate and fix the new mutation in ZEBOV in contrast to the past outbreaks. Meanwhile, considering the frequent occurrence of the mutation at 544 during the past outbreaks, the mutation at 82 might be beneficial for the Makona strain. To assess the influence of the A82V mutation in other ZEBOV strains, we measured the virus infectivity mediated by GP of prototype strain Mayinga. In contrast to the result from the Makona strain, in the Mayinga strain, the A82V mutation decreased the infectivity of the pseudotype VSVs to the cells (Fig. S6 in Supporting information). This suggests that the A82V mutation was beneficial only for Makona, but not for Mayinga. Amino acids differ by approximately 3% between Makona and Mayinga GPs. This might generate the difference in structures surrounding the residue 82, which is important for the viral entry to the host cell, and may affect a specific combination for beneficial

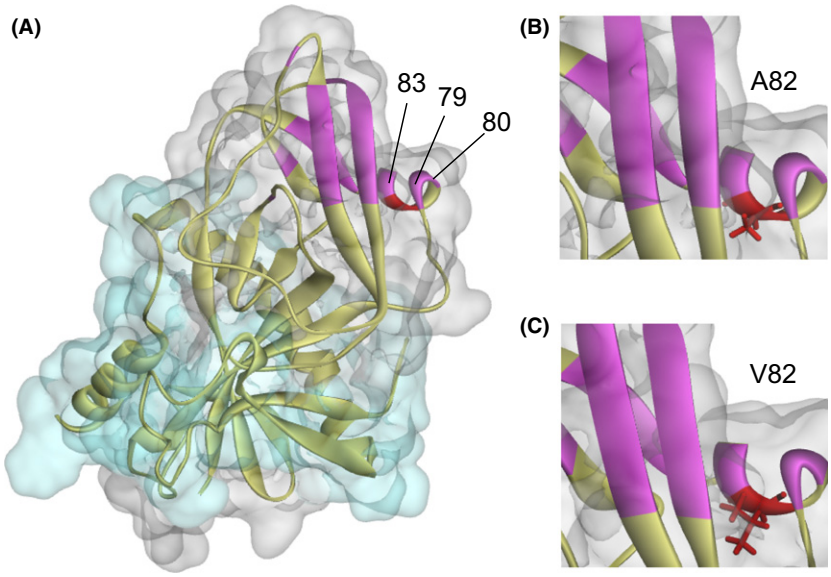

Figure 5 A82V is located near the NPC1 binding region. A: GP residues interacting with NPC1 are shown in magenta. In particular, residues near the position 82 are indicated. The residue at 82 is in red. B: Enlarged images for residues around 82 are on the left. The residues at 82 are B: alanine and C: valine.

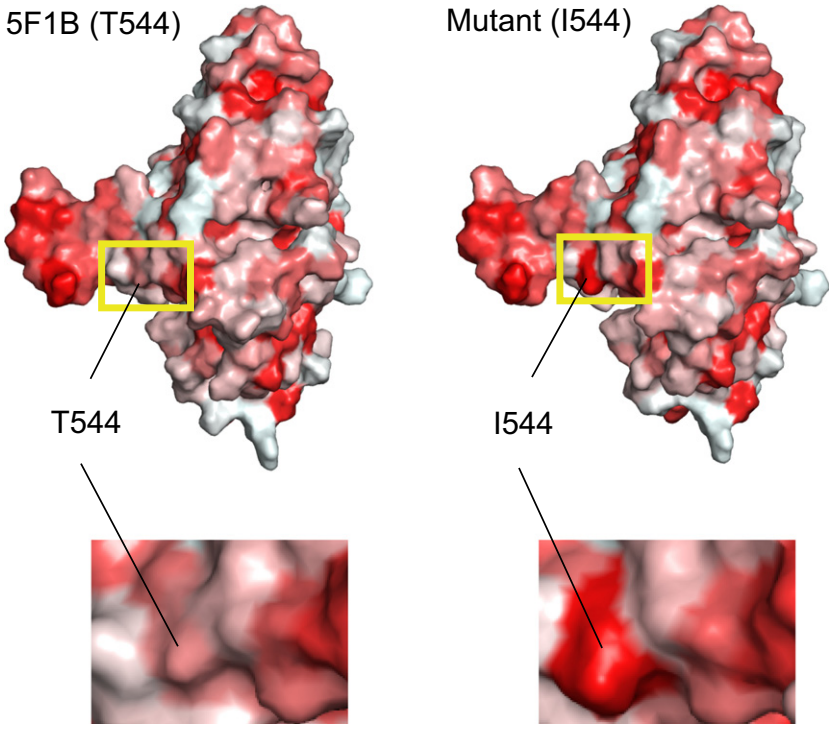

Figure 6 Hydrophobicity around residue 544 was increased by T544I mutation. The surface models of EBOV representing hydrophobicity are shown (hydrophilic surface is white and hydrophobic surface is red). The surfaces surrounding the 544 residue in yellow boxes are enlarged in the lower panels. The structure data were visualized by PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC).

mutations in each strain as discussed in Urbanowicz *et al.* (2016). An interesting observation was that the number of cases in West African countries dramatically increased after the A82V mutation was fixed in the population. There are many factors affecting this increase in these cases, such as mutations in other genomic regions and population density. However, the fixation of V82 may contribute to the spread of ZEBOV infection.

The T544I mutant has been frequently detected during the previous four ZEBOV epidemics, whereas the proportions of the mutant during each epidemic were low (Fig. S2 in Supporting information). This does not deny the occurrence of neutral genetic drift at this position. However, our structural analysis showed that the residue I544 stabilized the primed GP structure (mutation energy: −1.47 kcal/mol) by increasing the hydrophobicity around residue 544 (Fig. 6). This insight is further supported by the fact that the residue at 544 of Mayinga GP consists of a scaffold to form the hydrophobic surface of IFL, which is critical for viral-host membrane fusion in the late endosome (Gregory *et al.* 2011). An alanine mutation at 544 (I544A) decreased the IFL activity, due to the spread of residues forming a hydrophobic surface (Gregory *et al.* 2014). Consistent with the previous study (Gregory *et al.* 2014), we confirmed that the infectivity was dramatically decreased by the mutation to alanine at 544 in both strains Makona (T544A) and Mayinga (I544A) (Figs 4 and S6 in Supporting information). We found that there was a 4.3-fold increase in infectivity by the T544I mutation compared with the original Makona strain, whereas the increase in A82V infectivity was 1.8-fold (Fig. 4, upper panel). The difference in the increase in infectivity by the A82V and T544I mutations appears marginal; however, given that this infectivity is for single-round infection, viruses increase exponentially during the actual infection process in host cells, indicating that the mutation at 544 might be more beneficial for the spread of viral infection. However, in actual fact, this larger increase in infectivity at 544 appears to adversely affect the spread of viral infection. In fact, the mutation was not fixed in the population. The mutation may lead to the higher death rate and/or shorter incubation periods in patients,

and more rapid death after infection than those of the original Makona strain. These factors might reduce the frequency of contact between ZEBOV-infected cases and uninfected people if ZEBOV virulence is assumed to be largely determined by infectivity. Although the actual relationship between virulence and infectivity in humans remains unclear, we note that the difference in the infectivity between Makona with T544 and prototype Mayinga with I544 showed a similar tendency to those in the virulence. The original Makona strain (82A/544T) requires roughly 2 days longer to cause terminal diseases in cynomolgus macaque compared with Mayinga (544I) (Marzi *et al.* 2015). A peptide sequence in mouse-adapted GP containing amino acid at 544 was identified as a protective epitope, and induced CD4+ and CD8+ T cell responses in mice vaccinated with Venezuelan equine encephalitis virus replicons (Olinger *et al.* 2005). Therefore, there is another possibility that T544I mutation may also enhance recognition of virus infection by the host immune system. T544I mutation may have both advantages (increase infectivity) and disadvantages (enhance recognition by the host immune system) for ZEBOV in humans.

Considered together, we interpreted the transmission history of these mutants as evidence that the increase in infectivity induced by A82V was more moderate than that by T544I; thus, only the former could spread widely in the population. Currently, we have no effective treatments against Ebola virus disease, and EBOV outbreaks remain unpredictable. The amino acid changes in ZEBOV tended to disappear between epidemic waves (Azarian *et al.* 2015). However, certain nascent mutations that increase infectivity may re-emerge again through selective pressure in future Ebola virus outbreaks. Our findings from the current collaborative approach will provide insights into how ZEBOV evolved via GP, which helps our further understanding of potentially causative mutations that increase viral infectivity, virulence and transmissibility.

Experimental procedures

Sequences

All 1232 GP nucleotide sequences from complete ZEBOV genomes were obtained from GenBank (http://www.ncbi.nlm.nih.gov) on 19 January 2016. The ZEBOV GP sequence requires eight consecutive adenine residues at the RNA editing site to produce the full-length protein (Volchkov *et al.* 1995; Sanchez *et al.* 1996). Therefore, we inserted an adenine residue into the sequence (with seven consecutive adenines) when the nucleotide sequence was not a full-length GP alignment. If two or more GP sequences were identical, the sequence from the strain with the oldest sampling year was chosen. As a result, we extracted 147 representative GP sequences that were used as the core set for the current study (Table S1 in Supporting information).

Construction of a phylogenetic tree

For GP trees, GP representative nucleotide sequences were aligned by MAFFT with the L-INS-i option and default settings. A phylogenetic tree was then constructed using RAxML (Stamatakis 2006) with rapid 500 bootstrap replicates under the model (general time reversible model + gamma distribution) for the nucleotide alignment (Lanave *et al.* 1984). This phylogenetic tree was used for further analyses using PAML. For whole-genome trees, 805 ZEBOV genomes without undetermined base were used. The construction of phylogenetic trees was achieved using the same procedure to that of GP with rapid 100 bootstrap replications.

Positive selection analysis

The CODEML within the PAML 4.8 software package (Yang 2007) was used to estimate the sites under positive selection within the GP gene. We compared likelihoods of site models M1a vs. M2a and M7 vs. M8 to test for positive selection using the χ^2 distribution [likelihood ratio test (LRT); $df = 2$]. When LRTs indicated positive selection ($P \leq 0.01$), we examined the posterior probability [Bayes empirical Bayes (BEB) for sites outside MLD] and determined that they were under positive selection with a BEB of >0.95.

Structural analysis of ZEBOV GP

Crystal structures of ZEBOV glycoproteins, namely 3CSY (Lee *et al.* 2008), 5HJ3 (Bornholdt *et al.* 2016) and 5F1B (Wang *et al.* 2016), were optimized for inserting missing atoms in incomplete residues, modeling missing loop regions and deleting alternative conformations by the Prepare Protein protocol in Discovery Studio 4.1. The structural models of A82V and T544I mutants were then built based on optimized models using the Build Mutants protocol of the Macromolecular tool in Discovery Studio 4.1. To optimize the conformation of both the mutated and neighboring residues surrounding 82 and 544, the cut radius (5 Å) was used in this simulation because the mutated residues have a different size from the starting residue. The solvent-accessible surface area was calculated using the AREAIMOL program supported in the CCP4 program suite (Lee & Richards 1971; Saff & Kuijlaars 1997). Hydrophobicity was represented by the calculation of aggregation scores in Discovery Studio 4.1.

Cells and plasmids

Human embryonic kidney 293T, Huh-7, HeLa, SW13 and A549 cell lines were grown in Dulbecco's modified Eagle's medium (DMEM; Sigma, St Louis, MO, USA), supplemented with 10% fetal bovine serum (FBS; Sigma), 100 U/mL penicillin and 100 mg/mL streptomycin (Gibco, Life Technologies, Carlsbad, CA, USA), and maintained at 37 °C in a 5% CO_2. The GP coding sequence of the strain EBOV/H.sapiens-wt/GIN/2014/Makona-Gueckedou-C15 (Genbank accession number: KJ660346) was synthesized and cloned into the sequencing vector pCR2.1 (Invitrogen, Carlsbad, CA, USA). The single adenine was inserted at the RNA editing sites with inverse PCR using the KOD Plus mutagenesis kit (Toyobo, Osaka, Japan). The coding sequence of transmembrane GP was subcloned into the expression vector pCXN2 (Niwa *et al.* 1991). The resulting plasmids were designated pC-MakonaGP. The plasmid expressing GP of ZEBOV strain Mayinga (pCEboZ-GP) was provided by Dr Yoshihiro Kawaoka of the University of Tokyo, Japan. Single amino acid mutants at the position of 82 and 544 for Mayinga or Makona GP were obtained by inverse PCR with the KOD mutagenesis kit (Toyobo). The sequences of each wild type and its mutant plasmids were confirmed using an ABI Prism 3130 Genetic Analyzer (Applied Biosystems).

Infection assay by pseudotyped VSV

Pseudotype VSVs bearing GP of Ebola virus were generated as described previously (Tani *et al.* 2014). A total of 3×10^5 of 293T cells (producer cells) were seeded on collagen-coated 24-well culture plates 8 h prior to transfecting the plasmids. A total of 0.5 μg pC-MakonaGP, pCEboZ-GP and their amino acid mutants, or pCAGGS were transfected to the cells with TransIT LT-1 reagent (Mirus, Madison, WI). After 16 h of incubation, the cells were infected with G-complemented (*G) VSVΔG/Luc (*G-VSVΔG/Luc) (Tani *et al.* 2010) at a multiplicity of infection of 0.1. The virus was adsorbed for 1 h at 37 °C and then extensively washed four times with serum-free DMEM. After 24 h of incubation in DMEM supplemented with 2% FBS at 37 °C, the culture supernatants containing pseudotype VSVs were centrifuged at 800 ***g*** for 5 min at 4 °C to remove cell debris and subsequently stored at −80 °C until use. A total of 6×10^4 of Huh-7 cells were cultured on a 96-well clear-bottom plate at 37 °C in 5% CO_2 overnight. The cells were inoculated with 30 μL of culture fluids of producer cells, which contained pseudotype VSVs bearing EBOV GP (*EBOV GP-VSVΔG). After 24 h of incubation, infectivity of the pseudotype virus mediated by EBOV GP was assessed by luciferase activity with the Steady-Glo luciferase assay system (Promega Corporation, Madison, WI). The relative light unit value of luciferase was measured and determined with a microplate reader TriStar LB941 (Berthord, Tokyo, Japan).

Immunoblotting

Pseudotype VSVs bearing EBOV GP generated as described above were pelleted through a 20% (wt/vol) sucrose cushion at 80 000 rpm for 1 h in a TLA110 rotor (Beckman Coulter, Tokyo, Japan). The pellets were resuspended in PBS. The pseudotype VSV producer 293T cells were harvested in lysis A buffer (Urata *et al.* 2006). The lysates were centrifuged at 13 000 ***g*** for 10 min at 4 °C to separate insoluble pellets, following which the supernatants were used as samples. Each sample, boiled in loading buffer, was subjected to 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). The proteins were transferred to a nitrocellulose membrane (GE Healthcare, Tokyo, Japan) and reacted with mouse monoclonal anti-GP antibody (Takada *et al.* 2001), anti-VSV-M [23H12] antibody (Kerafast, Boston, MA, USA) or anti-β-actin antibody (Sigma) and horse radish peroxidase-conjugated anti-mouse IgG secondary antibody (Sigma). The proteins were detected with ECL prime Western blotting detection reagent (GE) in an image analyzer LAS-3000 (Fuji Film, Tokyo, Japan).

Acknowledgements

We thank Drs Tadashi Imanishi and Akiko Kanamori of Tokai University for their useful comments. We also thank Dr Hideki Tani of National Institute of Infectious Diseases for providing the pseudotype VSV system and Dr N'Faly Magassouba of National Donka Hospital. This study was supported by MEXT-Supported Program for the Strategic Research Foundation at Private Universities (to M.T.U. and S.N.), 2015 Tokai University School of Medicine Research Aid (to S.N.), Joint Usage/Research Center program of Institute for Virus Research, Kyoto University (to S.N.), Japan Agency for Medical Research and Development (15fk-0108039h0002) (to J.Y.) and CREST, Japan Science and Technology Agency (to K.S.).

References

Azarian, T., Lo Presti, A., Giovanetti, M., Cella, E., Rife, B., Lai, A., Zehender, G., Ciccozzi, M. & Salemi, M. (2015) Impact of spatial dispersion, evolution, and selection on Ebola Zaire Virus epidemic waves. *Sci. Rep.* **5**, 10170.

Baize, S., Pannetier, D., Oestereich, L. *et al.* (2014) Emergence of Zaire Ebola virus disease in Guinea. *N. Engl. J. Med.* **371**, 1418–1425.

Bale, S., Dias, J.M., Fusco, M.L., Hashiguchi, T., Wong, A.C., Liu, T., Keuhne, A.I., Li, S., Woods, V.L., Chandran, K., Dye, J.M. & Saphire, E.O. (2012) Structural basis for differential neutralization of Ebolaviruses. *Viruses* **4**, 447–470.

Bornholdt, Z.A., Ndungo, E., Fusco, M.L., Bale, S., Flyak, A.I., Crowe, J.E., Chandran, K. & Saphirea, E.O. (2016) Host-primed Ebola virus GP exposes a hydrophobic NPC1 receptor-binding pocket, revealing a target for broadly neutralizing antibodies. *mBio* **7**, e02154–15.

Brecher, M., Schornberg, K.L., Delos, S.E., Fusco, M.L., Saphire, E.O. & White, J.M. (2012) Cathepsin cleavage potentiates the Ebola virus glycoprotein to undergo a subsequent fusion-relevant conformational change. *J. Virol.* **86**, 364–372.

Carette, J.E., Raaben, M., Wong, A.C., Herbert, A.S., Obernosterer, G., Mulherkar, N., Kuehne, A.I., Kranzusch, P.J., Griffin, A.M., Ruthel, G., Dal Cin, P., Dye, J.M., Whelan, S.P., Chandran, K. & Brummelkamp, T.R. (2011) Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343.

Carstea, E.D., Morris, J.A., Coleman, K.G. *et al.* (1997) Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* **277**, 228–231.

Chandran, K., Sullivan, N.J., Felbor, U., Whelan, S.P. & Cunningham, J.M. (2005) Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science* **308**, 1643–1645.

Côtè, M., Misasi, J., Ren, T., Bruchez, A., Lee, K., Filone, C.M., Hensley, L., Li, Q., Ory, D., Chandran, K. & Cunningham, J. (2011) Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature* **477**, 344–348.

Cruz, J.C., Sugii, S., Yu, C. & Chang, T.Y. (2000) Role of Niemann-Pick type C1 protein in intracellular trafficking of low density lipoprotein-derived cholesterol. *J. Biol. Chem.* **275**, 4013–4021.

Davies, J.P. & Ioannou, Y.A. (2000) Topological analysis of Niemann-Pick C1 protein reveals that the membrane orientation of the putative sterol-sensing domain is identical to those of 3-hydroxy-3-methylglutaryl-CoA reductase and sterol regulatory element binding protein cleavage-activating protein. *J. Biol. Chem.* **275**, 24367–24374.

Diehl, W.E., Lin, A.E., Grubaugh, N.D. *et al.* (2016) Ebola Virus glycoprotein with Increased Infectivity dominated the 2013–2016 epidemic. *Cell* **167**, 1088–1098.

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

Feldmann, H. & Geisbert, T.W. (2011) Ebola haemorrhagic fever. *Lancet* **377**, 849–862.

Feldmann, H., Klenk, H.D. & Sanchez, A. (1993) Molecular biology and evolution of filoviruses. *Arch. Virol. Suppl.* **7**, 81–100.

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

Gregory, S.M., Harada, E., Liang, B., Delos, S.E., White, J.M. & Tamm, L.K. (2011) Structure and function of the complete internal fusion loop from Ebolavirus glycoprotein 2. *Proc. Natl Acad. Sci. USA* **108**, 11211–11216.

Gregory, S.M., Larsson, P., Nelson, E.A., Kasson, P.M., White, J.M. & Tamm, L.K. (2014) Ebolavirus entry requires a compact hydrophobic fist at the tip of the fusion loop. *J. Virol.* **88**, 6636–6649.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

Kuhn, J.H., Andersen, K.G., Baize, S. *et al.* (2014) Nomenclature- and database- compatible names for the two Ebola virus variants that emerged in Guinea and the Democratic Republic of the Congo in 2014. *Viruses* **6**, 4760–4799.

Ladner, J.T., Wiley, M.R., Mate, S. *et al.* (2015) Evolution and Spread of Ebola virus in Liberia, 2014–2015. *Cell Host Microbe* **18**, 659–669.

Lanave, C., Preparata, G., Saccone, C. & Serio, G.A. (1984) New method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20**, 86–93.

Lee, B. & Richards, F. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

Lee, J.E., Fusco, M.L., Hessell, A.J., Oswald, W.B., Burton, D.R. & Saphire, E.O. (2008) Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**, 177–182.

Liu, S.Q., Deng, C.L., Yuan, Z.M., Rayner, S. & Zhang, B. (2015) Identifying the pattern of molecular evolution for Zaire ebolavirus in the 2014 outbreak in West Africa. *Infect. Genet. Evol.* **32**, 51–59.

Marzi, A., Feldmann, F., Hanley, P.W., Scott, D.P., Günther, S. & Feldmann, H. (2015) Delayed disease progression in cynomolgus macaques infected with Ebola virus Makona strain. *Emerg. Infect. Dis.* **21**, 1777–1183.

Miller, E.H., Obernosterer, G., Raaben, M. *et al.* (2012) Ebola virus entry requires the host-programmed recognition of an intracellular receptor. *EMBO J.* **31**, 1947–1960.

Niwa, H., Yamamura, K. & Miyazaki, J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199.

Olinger, G.G., Bailey, M.A., Dye, J.M., Bakken, R., Kuehne, A., Kondig, J., Wilson, J., Hogan, R.J. & Hart, M.K. (2005) Protective cytotoxic T-cell responses induced by venezuelan equine encephalitis virus replicons expressing Ebola virus proteins. *J. Virol.* **79**, 14189–14196.

Park, D.J., Dudas, G., Wohl, S. *et al.* (2015) Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell* **161**, 1516–1526.

Saff, E.B. & Kuijlaars, A.B.J. (1997) Distributing many points on a sphere. *Math. Intell.* **19**, 5–11.

Sanchez, A., Trappier, S.G., Mahy, B.W., Peters, C.J. & Nichol, S.T. (1996) The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl Acad. Sci. USA* **93**, 3602–3607.

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

Schornberg, K., Matsuyama, S., Kabsch, K., Delos, S., Bouton, A. & White, J. (2006) Role of endosomal cathepsins in

entry mediated by the Ebola virus glycoprotein. *J. Virol.* **80**, 4174–4178.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.

Takada, A., Watanabe, S., Okazaki, K., Kida, H. & Kawaoka, Y. (2001) Infectivity-enhancing antibodies to Ebola virus glycoprotein. *J. Virol.* **75**, 2324–2330.

Tani, H., Iha, K., Shimojima, M., Fukushi, S., Taniguchi, S., Yoshikawa, T., Kawaoka, Y., Nakasone, N., Ninomiya, H., Saijo, M. & Morikawa, S. (2014) Analysis of Lujo virus cell entry using pseudotype vesicular stomatitis virus. *J. Virol.* **88**, 7317–7330.

Tani, H., Shiokawa, M., Kaname, Y., Kambara, H., Mori, Y., Abe, T., Moriishi, K. & Matsuura, Y. (2010) Involvement of ceramide in the propagation of Japanese encephalitis virus. *J. Virol.* **84**, 2798–2807.

Urata, S., Noda, T., Kawaoka, Y., Yokosawa, H. & Yasuda, J. (2006) Cellular factors required for Lassa virus budding. *J. Virol.* **80**, 4191–4195.

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

Volchkov, V.E., Becker, S., Volchkova, V.A., Ternovoj, V.A., Kotov, A.N., Netesov, S.V. & Klenk, H.D. (1995) GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214**, 421–430.

Volchkov, V.E., Feldmann, H., Volchkova, V.A. & Klenk, H.D. (1998) Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proc. Natl Acad. Sci. USA* **95**, 5762–5767.

Wang, H., Shi, Y., Song, J., Qi, J., Lu, G., Yan, J. & Gao, G.F. (2016) Ebola viral glycoprotein bound to its endosomal receptor Niemann-Pick C1. *Cell* **164**, 258–268.

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

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

Received: 4 November 2016
Accepted: 29 November 2016

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 ML tree of GP nucleotide sequences obtained from all ZEBOV genomes examined.

Figure S2 Percentages of ancestral and mutated residues at 82 and 544 of GP during each outbreak year.

Figure S3 Whole-genome phylogenetic tree from the 2014–2015 outbreak (A82V mutant).

Figure S4 Whole-genome phylogenetic tree from the 2014–2015 outbreak (T544I mutant).

Figure S5 Impacts of mutations at position 82 and 544 of Makona GP on entry into different cell types.

Figure S6 Infectivity of pseudotyped VSVs mediated by EBOV Mayinga GP and its mutants.

Table S1 Likelihood ratio test of M1-M2 and M7-M8 of PAML

Table S2 Positively selected sites for M2 and M8 of PAML

Table S3 Positively selected sites using other models in HyPhy