Evolutionary history of Ebola virus

Y. H. LI[1] AND S. P. CHEN[2]*

[1] *Center of Hematopoietic Stem Cell Transplantation, Affiliated Hospital of the Academy of Military Medical Sciences, Beijing, PR China*
[2] *Department of Laboratory Medicine, Affiliated Hospital of the Academy of Military Medical Sciences, Beijing, PR China*

SUMMARY

Since Ebola virus was discovered in 1970s, the virus has persisted in Africa and sporadic fatal outbreaks in humans and non-human primates have been reported. However, the evolutionary history of Ebola virus remains unclear. In this study, 27 Ebola virus strains with complete glycoprotein genes, including five species (Zaire, Sudan, Reston, Tai Forest, Bundibugyo), were analysed. Here, we propose a hypothesis of the evolutionary history of Ebola virus which will be helpful to investigate the molecular evolution of these viruses.

Key word: Molecular epidemiology.

INTRODUCTION

The family Filoviridae consists of two genera, Ebola virus (EBOV) and Marburg virus. According to the significant differences in the antigenicity and the nucleotide sequences, EBOV is subdivided into five species: Zaire (EBOV-Z), Sudan (EBOV-S), Reston (EBOV-R), Tai Forest (EBOV-TF, which was also known as Cote d'Ivoire ebola virus until 2010), and Bundibugyo (EBOV-B) [1, 2]. EBOV-S and EBOV-Z, which are the predominant EBOVs associated with known outbreaks, are more pathogenic than EBOV-R and EBOV-TF [3]. EBOV-TF has only caused a single non-fatal human infection, but EBOV-R has caused fatal infection in non-human primates [4]. However, EBOV-S, EBOV-Z, and EBOV-B often cause severe haemorrhagic diseases with markedly high case fatality rates (40–90%) [2, 5]. Due to the high biohazard risk, EBOV is classified as a BSL4 (biosafety level 4) agent based on high mortality rate, person-to-person transmission, potential aerosol infectivity, and absence of vaccines and therapies. Safe manipulation of EBOV requires maximum containment facilities.

The genome of EBOV is a single non-segmented, negative-stranded RNA (18·9 kb in length) with the following gene order: 3′ leader nucleoprotein (NP) – virion protein (VP) 35 – VP40 – glycoprotein (GP) – VP30 – VP24 – polymerase (L) – 5′ trailer. The GP differences between any two species range from 37% to 41% at the nucleotide level and from 34% to 43% at the amino-acid level [1]. However, variations within EBOV-Z species are very low (∼2–3%) [1]. Thus, GP nucleotides are usually used in the phylogenetic analysis of EBOV [6–9].

The first three known outbreaks of EBOV occurred during the 1970s in the Democratic Republic of the Congo (DRC) and Sudan [10–12]. No further cases were confirmed in Africa until late 1994. Since then, EBOV (EBOV-Z, EBOV-S, EBOV-TF, EBOV-B)

* Author for correspondence: Dr S. P. Chen, Department of Laboratory Medicine, Affiliated Hospital of the Academy of Military Medical Sciences, No. 8, Dongdajie Street, Fengtai District, Beijing 100071, PR China.
(Email: shpchen@hotmail.com)

Table 1. *Ebola viruses used in the study*

Strain name	Country	Isolation date	GenBank no.
Mayinga	DRC	1995	JQ352763
Mayinga	DRC	1976	NC_002549
Zaire 1995	DRC	1995	AY354458
034-KS	DRC	2008	HQ613402
Bouee-96	Gabon	1996	AY058898
Eckron-76	Gabon	1976	U81161
Gabon-94	Gabon	1994	U77384
GOR Ekata	Gabon	2001	EU051632
Mandza	Gabon	2003	EU051635
Mayibout	Gabon	1996	HQ849547
Siena (1992)	Italy	1992	U23417
Ivory Coast	Ivory Coast	1994	U28006
Philippine 1992	Philippines	1992	U23416
Reston08-A	Philippines	2008	FJ621583
Reston08-C	Philippines	2008	FJ621584
Reston08-E	Philippines	2008	FJ621585
Reston	Philippines	1996	AB050936
CH Lossi	RC†	2003	EU051633
Etoumbi	RC†	2005	EU051634
GOR1 Lossi	RC†	2002	EU051630
GOR2 Lossi	RC†	2002	EU051631
Boneface	Sudan	1976	FJ968794
Maleo	Sudan	1979	U23069
Yambio	Sudan	2004	EU338380
Gulu	Uganda	2000	AY344234
No name*	Uganda	2007	FJ217161
Pennsylvania	USA	1989	NC_004161

DRC, Democratic Republic of the Congo; RC, Republic of the Congo.

* No name: the strain was not named.

† The place of isolation was corrected according to Wittmann *et al.* [9].

has circulated in several African countries, including the Ivory Coast, DRC, Uganda, Republic of the Congo (RC), and Gabon (www.who.int). EBOV-R first emerged in the USA in 1989 from monkeys imported from the Philippines [13]. The subsequent outbreaks in the USA in 1990, Italy in 1992, and the Philippines in 1996 are all traced back to the Philippines [14, 15].

Swine and monkeys are hosts of EBOV-R [13, 16]. Chimpanzees, gorillas, and humans are also well-known as hosts of EBOV-Z [9]. With regard to reservoirs, fruit bats thus far are confirmed as reservoirs of EBOV-Z [17, 18]. Previous analysis of GP, NP, and L genes of EBOV-Z suggests that all viruses have recent common ancestries regardless of the sampling dates [9, 19, 20], while EBOV is estimated to be at least 1000–2100 years old [7]. These results,

at least at first sight, appear to contradict each other. One explanation for these results is that EBOV-Z experienced a recent genetic bottleneck [19]. However, it is unclear whether this explanation is viable or not.

To address this key question, EBOV strains with complete GP sequences, including five species, were analysed in this study. We found that since 1976 all EBOV-Z, EBOV-S, and EBOV-R strains traced back to around 1970, just at the time when the genetic diversity of EBOV declined to the lowest during its evolution history. Our analysis showed that EBOV experienced a recent genetic bottleneck.

METHODS

Sequences

All EBOVs with complete GP sequences were obtained as of 3 August 2012 from GenBank (www.ncbi.nlm.nih.gov). All viral strains had a known date and place of isolation. To facilitate analysis, all GP sequences used in this study were mRNA sequences which were proofread by RNA editing. For those strains having 100% similarity, only one strain was selected to remain in the dataset. Sequences containing stop codons were excluded. In total, there were 27 strains with a time span from 1976 to 2008 (Table 1). All the following analysis is based on the complete ORF sequences of GP. No recombinant viruses were identified in the cohort.

Maximum-likelihood (ML) tree construction

The nucleotide sequences of GP were aligned by ClustalW implemented in MEGA4 [21]. The ML phylogenetic tree was inferred by using the TREE-PUZZLE program [22]. The tree was rooted by using the complete GP sequence of Marburg virus (AF005734).

Molecular evolution analysis

The rate of nucleotide substitution per site and the time to the most recent common ancestor (TMRCA) were estimated by using the Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in BEAST v. 1.7.5 package (http://beast.bio.ed.ac.uk) [23]. We employed both strict and relaxed (uncorrelated exponential, uncorrelated lognormal) molecular clocks with different demographic models (constant size, exponential growth, logistic growth, expansion growth).

Model comparisons were done by calculation of Bayes factors based on the relative marginal likelihoods and posteriors of the models. The GTR+I +R model, which was selected using Akaike's Information Criterion (AIC) as implemented in jModelTest v. 2.1.1 [24], was found to be the best-fit nucleotide substitution model for our dataset. The population dynamics of EBOV was inferred using the Bayesian skyline plot model and the relaxed uncorrelated lognormal clock model of substitution. The TMRCA and treeModel.rootHeight parameters had prior distributions via the selected tree prior. The prior distributions for ac, ag, at, cg, and gt parameters were specified as gamma. The prior distribution for the constant.popSize parameter was specified as $1/x$. The prior distributions for other parameters were specified as uniform. Final chain length of at least 70 million was employed to make an effective sample size for parameters estimates >200. The resulting convergence was analysed using Tracer 1.5 (http://evolve.zoo.ox.ac.uk). To reveal the uncertainty in the estimations, the 95% high probability density (HPD) intervals in each case were also determined.

Selection pressures

Overall selection pressures acting on GP were determined as the ratio of non-synonymous (dN) to synonymous (dS) substitutions (dN/dS) per site using the pairwise method of Nei & Gojobori as implemented in MEGA4 [21]. To identify the positive selection sites in each species, Datamonkey [25] was employed and the single likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), internal FEL (IFEL), and random-effects likelihood (REL) methods were used (http://www.datamonkey.org).

RESULTS

Phylogenetic relationship

As shown in Figure 1, EBOV-S included strains from Uganda and Sudan; EBOV-Z included strains from DRC, RC and Gabon; EBOV-TF included the strain from the Ivory Coast; EBOV-B included the strain from Uganda; EBOV-R included strains from the Philippines, USA, and Italy. Regarding EBOV-Z, it was subdivided into two distinct lineages (lineage 1, lineage 2), of which lineage 1 included strains from Gabon/RC and lineage 2 included strains from Gabon/DRC.

Molecular clock analysis

To estimate the substitution rates and the TMRCA of EBOV, the best fit model for 27 strains with complete GP sequences (Table 1) were first analysed. Of all three molecular clock models (strict, relaxed uncorrelated exponential, relaxed uncorrelated lognormal), the relaxed molecular clocks performed better than the strict clock model (Table 2, Bayes factor >50). The logistic and exponential growth demographic model did not converge under the strict and relaxed (uncorrelated exponential and uncorrelated lognormal) molecular clock models. As shown in Table 2, there were no significant differences in the relative marginal likelihoods when the two molecular clock models (relaxed uncorrelated exponential, relaxed uncorrelated lognormal) and the two population models (constant size, expansion growth) were employed (Bayes factor <50). However, the Bayes factors based on the posteriors gave stronger support to the constant size demographic model under the relaxed uncorrelated lognormal clock model than any other molecular clock model and population model (Bayes factor >50).

As shown in Table 3, MCMC analysis under this best-fit model revealed that the evolutionary rate of EBOV was $10 \cdot 93 \times 10^{-4}$ (95% HPD $0 \cdot 52 \times 10^{-4}$ to $24 \cdot 61 \times 10^{-4}$) substitutions/site per year. The rates of three main species were as follow: EBOV-Z ($7 \cdot 66 \times 10^{-4}$, 95% HPD $3 \cdot 68 \times 10^{-4}$ to $11 \cdot 79 \times 10^{-4}$), EBOV-S ($13 \cdot 94 \times 10^{-4}$, 95% HPD $6 \cdot 06 \times 10^{-4}$ to $20 \cdot 64 \times 10^{-4}$), and EBOV-R ($10 \cdot 61 \times 10^{-4}$, 95% HPD $5 \cdot 08 \times 10^{-4}$ to $15 \cdot 88 \times 10^{-4}$) substitutions/site per year.

The TMRCA of EBOV since 1976 was estimated to occur in 751 (95% HPD 1320 B.C.–A.D. 1872). The TMRCA of EBOV-Z, EBOV-S, and EBOV-R were estimated to occur in 1971 (95% HPD 1960–1976), 1969 (95% HPD 1956–1976), and 1970 (95% HPD 1948–1987), respectively.

Genetic diversity

Figure 2a shows that the genetic diversity of EBOV remained constant before ~1900, and henceforth declined sharply until it reached its lowest around 1970. However, there was no significant difference in the genetic diversity for each EBOV species since ~1970. Specifically, the genetic diversity of EBOV-Z and EBOV-R increased very slightly at the beginning followed by stationary phases (Fig. 2b, c); the genetic diversity of EBOV-S remained constant all of the time (Fig. 2d).

Table 2. *Parameter estimates for different molecular clock models and population models*

Models	Likelihood (95% HPD)	Posterior (95% HPD)
Strict, constant size	−12710·9 (−12720·4 to −12702·4)	−12919·4 (−12937·5 to −12903·8)
Strict, expansion growth	−12710·7 (−12720·2 to −12702·2)	−12938·5 (−12963·1 to −12915·1)
Uncorrelated exponential, constant size	−12674·9 (−12685·8 to −12664·6)	−12866·6 (−12899·0 to −12839·6)
Uncorrelated exponential, expansion growth	−12674·9 (−12685·9 to −12664·6)	−12882·4 (−12917·7 to −12849·9)
Uncorrelated lognormal, constant size	**−12675·6 (−12686·5 to −12665·3)**	**−12864·1 (−12903·7 to −12829·0)**
Uncorrelated lognormal, expansion growth	−12675·8 (−12686·7 to −12665·3)	−12882·9 (−12922·3 to −12847·0)

HPD, High probability density.
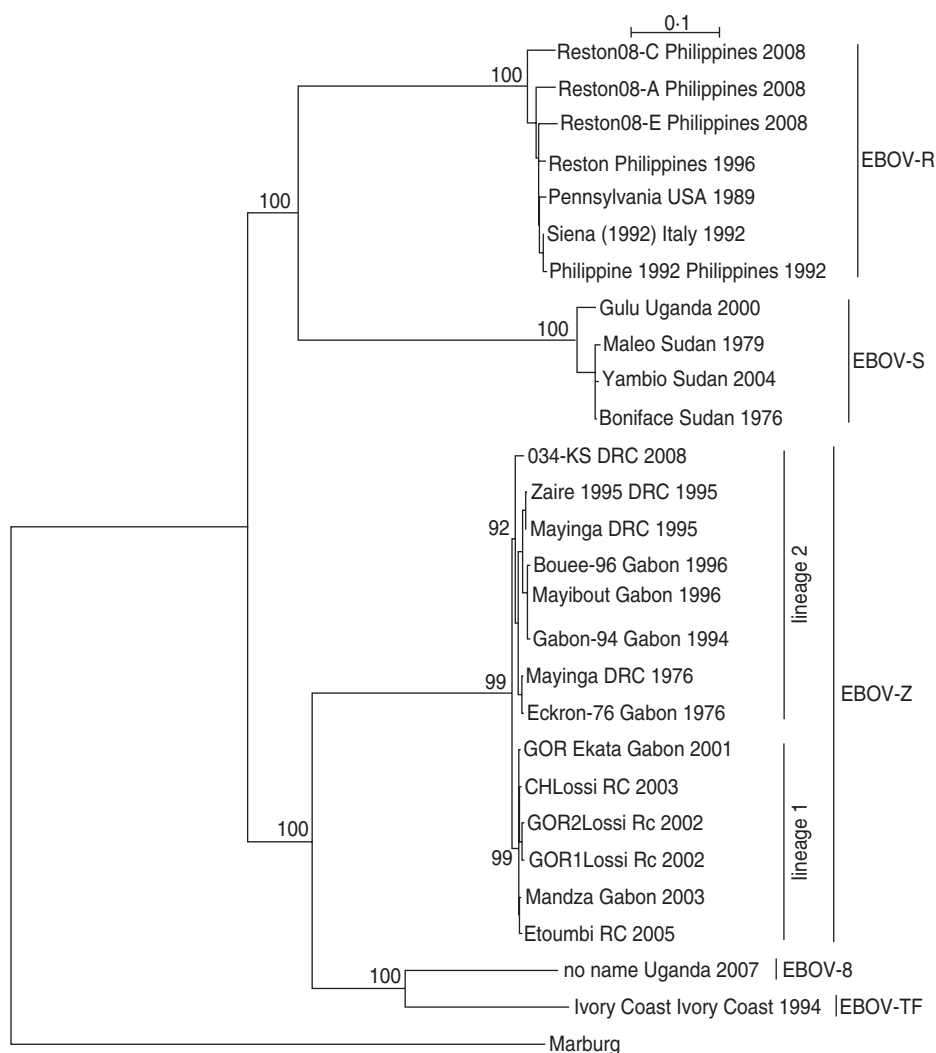The best fit model appears in bold.

Fig. 1. The phylogenetic relationship of Ebola virus (EBOV) isolates. The complete GP nucleotide sequences were analysed by the maximum-likelihood method using the TREE-PUZZLE program. Taxon names corresponded to 'strain name + isolation place + isolation time'. Bootstrap values are shown on the key nodes of the trees. The trees were rooted by Marburg virus (AF005734).

EBOV under positive selection

To assess the selection pressures acting on GP, the average dN/dS value measured by MEGA4 is shown in Table 4. The dN/dS values of three species (EBOV-S, EBOV-Z, EBOV-R) and the dataset including all five species were located between 0·229 and 0·38, which indicated that EBOV was under purifying selection.

Table 3. *Evolutionary processes of Ebola virus based on glycoprotein*

Subtype	Evolutionary rates (10^{-4} substitution rate, substitutions/ site per year) (95% HPD)	TMRCA (95% HPD)
EBOV-Z	7·66 (3·68–11·79)	1971 (1960–1976)
EBOV-S	13·94 (6·06–20·64)	1969 (1956–1976)
EBOV-R	10·61 (5·08–15·88)	1970 (1948–1987)
Total	10·93 (0·52–24·61)	751 (1320 B.C.– A.D. 1872)

TMRCA, Time to most recent common ancestor; HPD, high probability density.
EBOV-TF and EBOV-B were not analysed due to limited sequences.

Table 4. *Selection pressure acting on the glycoprotein of Ebola virus*

Subtype	dS	dN	dN/dS
EBOV-Z	0·011	0·048	0·229
EBOV-S	0·021	0·064	0·328
EBOV-R	0·021	0·064	0·329
Total	0·278	0·732	0·380

dS and dN are the number of synonymous and non-synonymous substitutions per site.
EBOV-TF and EBOV-B were not analysed due to limited sequences.

Meanwhile, selection pressure analysis (Table 5) was also performed by using the online server Datamonkey. Sites were considered to be under positive selection if at least two of the methods (SLAC, FEL, IFEL, REL) indicated this with high statistical significance ($P < 0.1$ or Bayes factor >50). Several positive selection sites were identified (Table 5). For instance, two sites (377, 443) were identified as being under strong positive selection using two different detection methods (IFEL, REL) in EBOV-Z; one site (229) was identified as being under strong positive selection using three different detection methods (FEL, IFEL, REL) in EBOV-R. With regard to EBOV-S, only one potential site (503) was identified as being under positive selection using one detection method (REL).

DISCUSSION

The polymerase is one of the most conserved proteins of EBOV, and GP is the least conserved [26]. Although GP is always used in evolutionary analysis of EBOV [7, 9, 19, 20], whether it is a better protein for such analysis than others is currently unclear. Two previous studies [19, 20] demonstrated that there were no significant differences in the evolutionary rates between the GP and L genes of EBOV-Z (GP ≈ 8.0×10^{-4} substitutions/site per year, L = 1.1×10^{-3} substitutions/site per year) as confidence intervals overlapped, which suggested that the least conserved GP could produce evolutionary rates similar to the most conserved polymerase (L). To further validate this deduction, we chose another species (EBOV-S) as a model to estimate the evolutionary rates based on GP and L in this study. Results showed that the rates based on GP and L were 13.94×10^{-4} (95% HPD 6.06×10^{-4} to 20.64×10^{-4}) substitutions/ site per year (Table 3) and 23.14×10^{-4} (95% HPD 16.49×10^{-4} to 31.66×10^{-4}) substitutions/site per year, respectively, which also showed no significant differences (Bayes factor <50). These aforementioned results clearly suggest that although GP is not the most conserved protein, it is still reliable for use in evolutionary analysis of EBOV. Of note is that the evolutionary rate of EBOV is 10.84×10^{-4} substitutions/site per year, and that there are no significant differences in the evolutionary rates between species (EBOV-Z, EBOV-S, EBOV-R) (Table 3).

With regard to the ancestor of EBOV, it was estimated to be 1000–2100 years old by analysing a dataset including four species (EBOV-Z, EBOV-S, EBOV-TF, EBOV-R) [7]. Although the newly identified species EBOV-B was included in the present study, the ancestor was 1257 years old (95% HPD 136–3328), which was also similar with Suzuki's result [7]. However, the ancestors of the three main species (EBOV-Z, EBOV-S, EBOV-R) emerged around 1970 (Table 3), which was ahead of the time that the viruses were first isolated (EBOV-Z and EBOV-S in 1976, EBOV-R in 1990). This raised the question 'Why have EBOVs been circulating for centuries, but only emerged recently?' Biek *et al.* proposed that EBOV-Z experienced a recent genetic bottleneck [19]. Here, we propose a hypothesis that the genus EBOV, not only the species EBOV-Z, also experienced a recent genetic bottleneck (Fig. 2a). Before EBOV emerged around 751 (95% HPD 1320 B.C.–A.D. 1872), the viruses had been circulating in small mammals (bats, rodents, shrews, tenrecs, marsupials, etc.) [27]. Although these animals (such as bats) were infected [28, 29], there was no evidence to show that they would die [28] which suggested a

Table 5. *Positive selection analysis based on the glycoprotein of Ebola virus*

Subtype	SLAC	FEL	IFEL	REL
EBOV-Z	–	–	**377**, **443**	331, 368, **377**, 389, 430, **443**, 483, 544
EBOV-S	–	–	–	503
EBOV-R	–	**229**	178, **229**, 430, 456, 467, 482	**229**

SLAC, Single likelihood ancestor counting; FEL, fixed effects likelihood; IFEL, internal FEL; REL, random effects likelihood.

Bold values denote amino-acid sites under positive selection by more than one method.

–, No positive selection sites were identified.

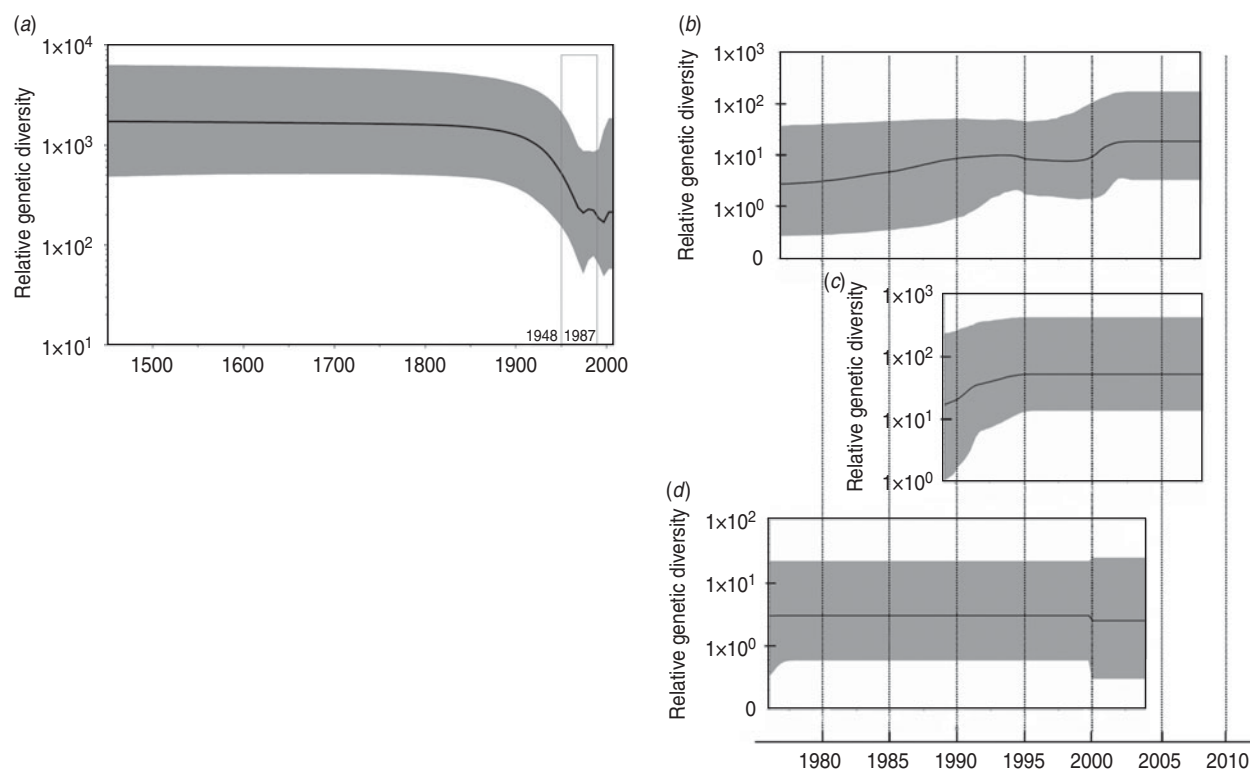EBOV-TF and EBOV-B were not analysed due to limited sequences.

Fig. 2. Relative genetic diversities of Ebola virus (EBOV) and each species over time. (*a*) EBOV; (*b*) EBOV-Z; (*c*) EBOV-R; (*d*) EBOV-S. The box [in panel (*a*)] denotes the time span from 1948 to 1987 which was the confidence intervals covering EBOV-Z, EBOV-S, and EBOV-R.

balance between these reservoirs and EBOV. However, this balance was broken around 1900 which was characterized by a rapid drop in genetic diversities of EBOV (Fig. 2*a*). During this process, most lineages of each species became extinct due to many factors, such as climate change, human activities, a sharp decrease in the numbers of reservoir animals or other possibilities. However, probably due to positive selection (Table 5) on GP which is well-known to be involved in receptor binding and fusion with cellular membranes, few lineages which obtained broader tropism and higher fitness thus had the ability

to infect primates around 1970 by direct exposure [30]. The similar examples were avian virus H5N1 and H7N9, which could now cross the species barrier to infect humans [31, 32]. Because there were no significant differences in the genetic diversities of EBOV since 1970 (Fig. 2*a–d*), the surviving viruses might therefore become the sole lineages circulating in reservoirs and primates since then (Fig. 1). Specifically, EBOV-Z was documented as having the ability to move a long distance along with the migratory reservoirs (bats) and outbreaks caused by this species thus occurred at the front of an advancing wave [20].

EBOV-TF and EBOV-B were first isolated in 1994 and 2007, respectively [2, 33]. According to our hypothesis, we predict that EBOV-TF and EBOV-B are also likely to have emerged around 1970, just like other three species (EBOV-Z, EBOV-S, EBOV-R). However, up to now there has only been one reported outbreak caused by each of these two species [2, 33], although EBOV-TF and EBOV-B could cause human infections. Here, we present three alternative explanations of what could cause a few outbreaks associated with EBOV-TF and EBOV-B. One possibility is that EBOV-TF and EBOV-B have lower pathogenicity in humans. As described previously, EBOV-TF only caused non-fatal infection in humans [33], and the case fatality associated with EBOV-B (36%) was much lower than that observed for EBOV-Z (80–90%) and EBOV-S (50–55%) [2]. Thus, those cases infected with mild manifestation might go unreported, which could lead to undiscovered outbreaks. The second possibility is that there are fewer opportunities for humans to be under direct exposure to the reservoirs of the two species. The infection caused by EBOV-B in 1994 was due to close direct contact with an infected chimpanzee [33], while the origin of EBOV-B, which caused the outbreak in 2007, remains unclear. This might be due to special transmission routes or fewer reservoir animals, the two species predominantly circulating in reservoirs, and humans (or other hosts) having fewer chances of becoming infected. The third possibility is that these viruses are not well adapted to humans. As described previously [30], there was an apparent putative transmission chain for EBOV-Z. With regard to EBOV-TF and EBOV-B, they might have had difficulty in circulating in humans because they were not well adapted to humans or there were low viral loads in most primary cases [2], which rarely caused subsequent human-to-human transmission. To address the above issues about EBOV-TF and EBOV-B, studies in epidemiological surveillance and pathogenicity differences in hosts need to be performed.

Adaptive evolution increases viral fitness and thus might play a role in the virus evolution, which was characterized by shifts in host tropism, immune pressure and cellular milieu [34–39]. With regard to EBOV described in this study, positive selection sites were also found in the viruses, especially in EBOV-Z and EBOV-R (Table 5). GP is well-known to be involved in receptor binding and fusion with cellular membranes. Thus, positive selection might play an important role in shaping EBOV, which increased viral fitness and facilitated the viruses to infect primates.

In summary, the evolutionary history of EBOV has been described in the present study. EBOV had been circulating in reservoir animals for centuries. Since ~1900, most viral lineages began to disappear due to a genetic bottleneck. Therefore only those few with broader tropism and higher fitness could survive to infect primates, which caused the outbreaks reported since 1976.

REFERENCES

1. **Sanchez A, et al.** The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proceedings of the National Academy of Sciences USA* 1996; **93**: 3602–3607.
2. **Towner JS, et al.** Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathogens* 2008; **4**: e1000212.
3. **Fisher-Hoch SP, et al.** Pathogenic potential of filoviruses: role of geographic origin of primate host and virus strain. *Journal of Infectious Diseases* 1992; **166**: 753–763.
4. **Feldmann H, et al.** Ebola haemorrhagic fever. *Lancet* 2011; **377**: 849–862.
5. **Sanchez A, et al.** Filoviridae: Marburg and Ebola viruses. In: Knipe DM, Howley PM, eds. *Fields Virology*. Philidelphia: Lippincott Williams and Williams, 2007, pp. 1409–1448.
6. **Georges-Courbot MC, et al.** Isolation and phylogenetic characterization of Ebola viruses causing different outbreaks in Gabon. *Emerging Infectious Diseases* 1997; **3**: 59–62.
7. **Suzuki Y, et al.** The origin and evolution of Ebola and Marburg viruses. *Molecular Biology and Evolution* 1997; **14**: 800–806.
8. **Volchkov V, et al.** Emergence of subtype Zaire Ebola virus in Gabon. *Virology* 1997; **232**: 139–144.
9. **Wittmann TJ, et al.** Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. *Proceedings of the National Academy of Sciences USA* 2007; **104**: 17123–17127.
10. **Baron RC, et al.** Ebola virus disease in southern Sudan: hospital dissemination and intrafamilial spread. *Bulletin of the World Health Organization* 1983; **61**: 997–1003.

11. **Johnson KM.** Ebola haemorrhagic fever in Zaire, 1976. *Bulletin of the World Health Organization* 1978; **56**: 271–293.

12. **Simpson DIH.** Ebola haemorrhagic fever in Sudan, 1976. Report of a WHO/International Study Team. *Bulletin of the World Health Organization* 1978; **56**: 247–270.

13. **Jahrling PB,** *et al.* Preliminary report: isolation of Ebola virus from monkeys imported to USA. *Lancet* 1990; **335**: 502–505.

14. **Miranda ME,** *et al.* Epidemiology of Ebola (subtype Reston) virus in the Philippines, 1996. *Journal of Infectious Diseases* 1996; **179**: 115–119.

15. **Rollin PE,** *et al.* Ebola (subtype Reston) virus among quarantined nonhuman primates recently imported from the Philippines to the United States. *Journal of Infectious Diseases* 1999; **179**: 108–114.

16. **Barrette RW,** *et al.* Discovery of swine as a host for the Reston ebolavirus. *Science* 2009; **325**: 204–206.

17. **Leroy EM,** *et al.* Fruit bats as reservoirs of Ebola virus. *Nature* 2005; **438**: 575–576.

18. **Swanepoel R,** *et al.* Experimental inoculation of plants and animals with Ebola virus. *Emerging Infectious Diseases* 1996; **2**: 321–325.

19. **Biek R,** *et al.* Recent common ancestry of Ebola Zaire virus found in a bat reservoir. *PLoS Pathogens* 2006; **2**: e90.

20. **Walsh PD,** *et al.* Wave-like spread of Ebola Zaire. *PLoS Biology* 2005; **3**: e371.

21. **Tamura K,** *et al.* MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 2007; **24**: 1596–1599.

22. **Schmidt HA,** *et al.* TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 2002; **18**: 502–504.

23. **Drummond AJ,** *et al.* BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 2007; **7**: 214.

24. **Darriba D,** *et al.* jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 2012; **9**: 772.

25. **Pond SL,** *et al.* Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005; **21**: 2531–2533.

26. **Sanchez A,** *et al.* Complete genome sequence of an Ebola virus (Sudan species) responsible for a 2000 outbreak of human disease in Uganda. *Virus Research* 2005; **113**: 16–25.

27. **Taylor DJ,** *et al.* Filoviruses are ancient and integrated into mammalian genomes. *BMC Evolutionary Biology* 2010; **10**: 193.

28. **Hayman DTS,** *et al.* Long-term survival of an urban fruit bat seropositive for Ebola and lagos bat viruses. *PLoS ONE* 2010; **5**: e1978.

29. **Hayman DTS,** *et al.* Ebola virus antibodies in fruit bats, Ghana, West Africa. *Emerging Infectious Diseases* 2012; **18**: 1207–1209.

30. **Leroy EM,** *et al.* Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Diseases* 2009; **9**: 723–727.

31. **Watanabe Y,** *et al.* The changing nature of avian influenza A virus (H5N1). *Trends in Microbiology* 2012; **20**: 11–20.

32. **Gao R,** *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *New England Journal of Medicine* 2013; **368**: 1888–1897.

33. **Le Guenno B,** *et al.* Isolation and partial characterization of a new strain of Ebola virus. *Lancet* 1995; **345**: 1271–1274.

34. **Tsetsarkin KA,** *et al.* Chikungunya virus: evolution and genetic determinants of emergence. *Current Opinion in Virology* 2011; **1**: 310–317.

35. **Sun M,** *et al.* Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China. *PLoS ONE* 2012; **7**: e36577.

36. **Leal E,** *et al.* Selective pressures of human immunodeficiencey virus type 1 (HIV-1) during pediatric infection. *Infection Genetics and Evolution* 2007; **7**: 694–707.

37. **Wei K,** *et al.* Evolution and adaptation of hemagglutinin gene of human H5N1 influenza virus. *Virus Genes* 2012; **44**: 450–458.

38. **Ping J,** *et al.* Genomic and protein structural maps of adaptive evolution of human influenza A virus to increased virulence in the mouse. *PLoS ONE* 2011; **6**: e21740.

39. **Botosso VF,** *et al.* Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. *PLoS Pathogens* 2009; **5**: e1000254.