REVIEW

doi:10.1038/nature19790

The evolution of Ebola virus: Insights from the 2013–2016 epidemic

Edward C. Holmes[1], Gytis Dudas[2,3], Andrew Rambaut[3,4,5] & Kristian G. Andersen[6,7,8]

The 2013–2016 epidemic of Ebola virus disease in West Africa was of unprecedented magnitude and changed our perspective on this lethal but sporadically emerging virus. This outbreak also marked the beginning of large-scale real-time molecular epidemiology. Here, we show how evolutionary analyses of Ebola virus genome sequences provided key insights into virus origins, evolution and spread during the epidemic. We provide basic scientists, epidemiologists, medical practitioners and other outbreak responders with an enhanced understanding of the utility and limitations of pathogen genomic sequencing. This will be crucially important in our attempts to track and control future infectious disease outbreaks.

The 2013–2016 Ebola virus disease (EVD) epidemic in West Africa appears to have begun following human contact with an animal (probably bat) reservoir of Ebola virus (EBOV) in December 2013, in the small village of Meliandou in Guéckédou Prefecture, Guinea[1]. After this initial spill-over infection, the outbreak remained undetected for several months and spread via chains of sustained human-to-human transmission, with no evidence of additional zoonotic transfers from the animal reservoir[1–4]. By the time that EBOV (a lineage later named the Makona variant[5]) was confirmed in March 2014, several villages, towns and larger cities had reported cases[1]. When the World Health Organization declared the EVD outbreak to constitute a Public Health Emergency of International Concern in August 2014[6], EBOV had already spread across country borders with more than a thousand cases reported in Guinea, Sierra Leone, Liberia and Nigeria. In the epidemic that followed, a total of 28,646 confirmed and suspected cases of EVD were documented, with 11,323 recorded deaths, making it by far the largest outbreak of EVD on record[7].

Ebola virus is a negative-sense single-strand RNA ((−)ssRNA) virus with a 19-kilobase genome and, like most other RNA viruses, quickly generates mutations through error-prone replication. Until recently, genomic studies of infectious disease outbreaks were necessarily retrospective, occurring after the pathogen had either been eradicated or developed endemic transmission in the host population[8–12]. However, recent developments in high-throughput next-generation sequencing (NGS)[13–16] enabled rapid and in-depth viral genomic surveillance during the 2013–2016 EVD epidemic[1–3,17–26]. Indeed, with the advent of NGS it is now possible to generate pathogen genomic data directly from diagnostic patient samples[2,3,17–27] within days or hours of the sample being taken[25,26], and in challenging field situations[19,23,25,26]. The resulting large-scale sequence data sets provide new opportunities for the epidemiological investigation of transmission chains and the improvement of outbreak responses[28]. In the case of the 2013–2016 EVD epidemic, the sequence data generated have revealed key aspects of the patterns and processes of EBOV evolution as the epidemic proceeded[2,3,20–22,24–26,29,30]. Hence, not only was the 2013–2016 epidemic a landmark in the epidemiological history of EBOV, but the size of the resulting genomic data set—over 1,500 full-length EBOV Makona sequences (Table 1), or approximately 5% of those infected—also makes it one of the most densely sampled infectious disease outbreaks (Fig. 1a, b). Although sequence data have been generated from outbreaks of viral disease for over 30 years[8–12,31,32], the sheer size of the

Table 1 | Overview of EBOV sequencing studies performed during the 2013–2016 epidemic

Study	Platform	Method	Sequencing location	Case location	No. of seqs
Baize, S. *et al.* (Apr. 2014)[1]	Sanger	Amplified	International	Guinea	3
Gire, S. K. *et al.* (Sep. 2014)[2]	Illumina	Direct	International	Sierra Leone	79
Hoenen, T. *et al.* (Apr. 2015)[17]	Sanger	Amplified	International	Mali	4
Bell, A. *et al.* (May. 2015)[18]	Illumina	Direct	International	UK	3
Park, D. J. *et al.* (Jun. 2015)[3]	Illumina	Direct	International	Sierra Leone	232
Kugelman, J. R. *et al.* (Jul. 2015)[19]	Illumina	Direct	In-country/ Liberia	Liberia	25
Simon-Loriere, E. *et al.* (Aug. 2015)[20]	Illumina	Direct	International	Guinea	85
Carroll, M. W. *et al.* (Aug. 2015)[21]	Illumina	Direct	International	Guinea/ Liberia	179
Tong, Y. G. *et al.* (Aug. 2015)[22]	BGISEQ-100	Amplified	?	Sierra Leone	175
Smits, S. L. *et al.* (Sep. 2015)[23]	Ion Torrent	Amplified	In-country/ Sierra Leone	Sierra Leone	49
Ladner, J.T. *et al.* (Dec. 2015)[24]	Illumina	Direct	International	Liberia	140
Quick, J. *et al.* (Feb. 2016)[25]	MinION	Amplified	In-country/ Guinea	Guinea	137
Hoenen, T. *et al.* (Feb. 2016)[93]	MinION	Amplified	In-country/ Liberia	Liberia	8
Arias, A. *et al.* (Jun. 2016)[26]	Ion Torrent	Amplified	In-country/ Sierra Leone	Sierra Leone	554

Summary of the different sequencing efforts performed during the 2013–2016 EVD epidemic, noting sequencing platform. We include the following parameters for each study: method (Direct, no PCR amplification and/or enrichment; Amplified, material amplified via amplicon-based PCR before viral sequencing), sequencing location, country of origin for the sequenced samples (case location) and number of EBOV genomes produced.

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and Sydney Medical School, Charles Perkins Centre, University of Sydney, Sydney, New South Wales 2006, Australia. [2]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [3]Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3FL, UK. [4]Centre for Immunology, Infection and Evolution, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3FL, UK. [5]Fogarty International Center, National Institutes of Health, MSC 2220 Bethesda, Maryland 20892, USA. [6]The Scripps Research Institute, Department of Immunology and Microbial Science, La Jolla, San Diego, California 92037, USA. [7]The Scripps Research Institute, Department of Integrative Structural and Computational Biology, La Jolla, San Diego, California 92037, USA. [8]Scripps Translational Science Institute, La Jolla, San Diego, California 92037, USA.

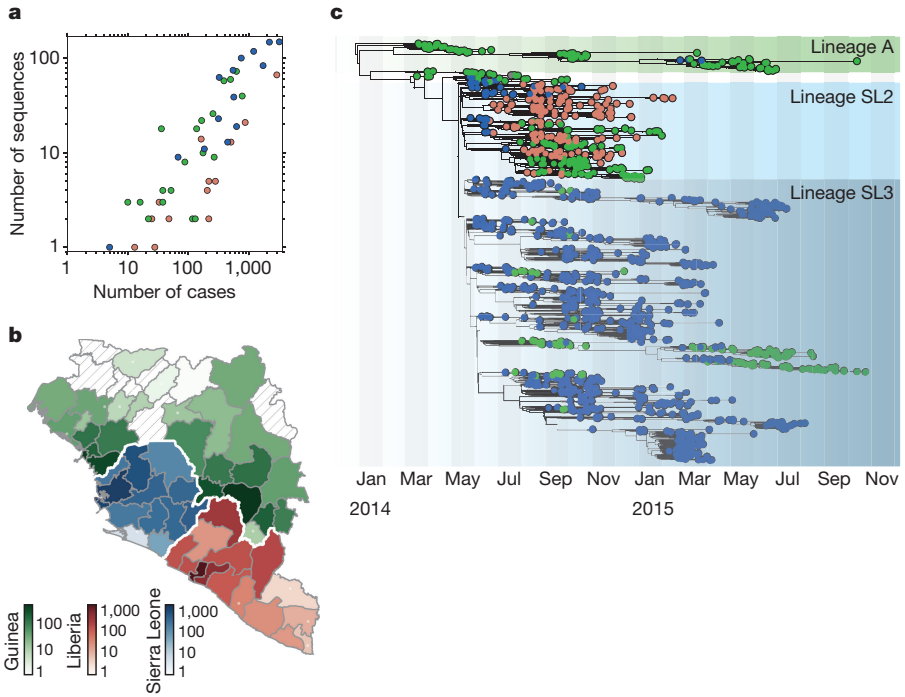


Figure 1 | Evolution of EBOV during the 2013–2016 outbreak showing the extent and location of virus sampling. a, Sampling during the 2013–2016 EVD epidemic. Sequencing efforts closely match confirmed and suspected case numbers in each administrative division of Guinea (green), Liberia (red) and Sierra Leone (blue) (Spearman correlation coefficient = 0.91). **b**, Map of the three countries most affected by EVD during the 2013–2016 EVD epidemic. Administrative divisions in Guinea, Liberia and Sierra Leone are shown in green, red and blue, respectively, and coloured according to the cumulative numbers of confirmed and suspected cases throughout the epidemic. Hatched areas indicate divisions that never reported any cases. The boundary data for the maps is from GADM (http://www.gadm.org). **c**, Temporal phylogeny of all publicly available EBOV genomes estimated using BEAST[97]. Three lineages identified in previous studies[2,21,25] are marked with coloured backgrounds. The sequence alignment was partitioned into four categories: codon positions 1, 2 and 3, and non-coding intergenic regions. Changes in each of the four partitions were modelled according to the HKY+Γ_4 nucleotide substitution model with relative rates between partitions. Tip dates were used to calibrate a relaxed molecular clock with rates drawn from a lognormal distribution[54] with an uninformative prior placed on the mean of the distribution. A flexible 'skygrid' tree prior was used to allow for changes in effective population sizes over time. Each tip is coloured according to the country where the patient was most likely to have been infected: green for Guinea, red for Liberia and blue for Sierra Leone. Data correct as of 19 April, 2016. *Nature* remains neutral with regard to jurisdictional claims in published maps.

data set, the widespread spatial coverage (Fig. 1a), and the contemporary nature of the EBOV data provide the first in-depth genomic anatomy of an epidemic, setting a benchmark for future outbreak responses. Here, we describe how pathogen sequences produced during the 2013–2016 EVD epidemic provided key insights into EBOV genomic epidemiology and molecular evolution, and note the lessons that need to be learned for the effective study of future outbreaks.

Ebola virus disease in humans

Ebola virus (species *Zaire ebolavirus*) is one of four viruses—with Sudan virus, Taï Forest virus, and Bundibugyo virus—within the genus *Ebolavirus* that cause severe disease in humans and other primates. The final member of the genus is Reston virus, although infection with this virus does not appear to cause human disease[33]. All ebolaviruses are members of the family *Filoviridae*, which also includes *Lloviu cuevavirus* (genus *Cuevavirus*) and the severe human pathogen Marburg virus (genus *Marburgvirus*). It is believed that bats serve as the primary reservoir for EBOV[34,35]. However, EBOV infections have been confirmed in only a small number of mammalian species and it is unclear whether the virus infects a wider range of animal hosts that have yet to be sampled. Some evidence for a broader host distribution, at least in the evolutionary past, comes from the observation that endogenous filoviruses are present in the genomes of diverse mammalian species, including marsupials[36,37].

EBOV in humans was first described in Zaire (now the Democratic Republic of the Congo (DRC)) in 1976, where, over a two-month period, it led to an outbreak of 318 cases with an 88% case-fatality rate (CFR)[38]. CFRs, however, are difficult to estimate for EVD[39,40], so such numbers should be interpreted with caution. Between 1977 and 2014, 12 smaller outbreaks were reported in Middle Africa, with 32–315 cases and CFRs ranging from 47% to 89%[41]. The 2013–2016 EVD epidemic is therefore notable not only for its duration and magnitude, but also as the first outbreak in West Africa and the first in which case exportations and nosocomial transmissions were reported outside of Africa[41]. However, despite the scale of the 2013–2016 EVD epidemic, infection with EBOV Makona appears to lead to similar disease characteristics and transmission profiles as previous EBOV outbreak variants[42,43]. For example, the CFR for the 2013–2016 EVD epidemic appears to be around 70%[1,39,40,43] and estimates for the basic reproduction number (R_0) fall between 1.5 and 2.5, both of which are comparable to calculations from previous outbreaks[39,44–47].

Origin of the 2013–2016 Ebola virus disease epidemic

Evolutionary analyses of genome sequences from the 2013–2016 EVD epidemic have provided a clear picture of the origin and spread of EBOV Makona[1–3,20–22,24–26,48]. One of the most important early questions was whether the epidemic was the result of a single cross-species transmission event into humans, or whether there were repeated zoonotic events from a widespread animal EBOV reservoir. Owing to the high genetic similarity of virus genomes sampled from the beginning of the epidemic, a single spill-over infection seems the more likely[1,2]. Phylogenetic analyses also make it clear that once the outbreak was established, later lineages of EBOV Makona had descended from those circulating earlier in the epidemic[2,3,20–22,24,25] (Fig. 1c). This is in contrast to some of the earlier EVD outbreaks, in which epidemiological and sequence-based investigations have provided evidence for multiple spill-over infections[49,50].

Sequence-based findings are consistent with epidemiological investigations into the timing of the 2013–2016 EVD epidemic, which placed the first case around late December 2013 in Guinea[1]. In agreement with this, molecular clock dating analyses suggest that the common ancestor of all sequenced EBOV Makona lineages be placed at the beginning of 2014[2,21,48], with lineages in Guinea falling close to the root of the tree (Fig. 1c). These studies also showed that EBOV Makona diverged from other EBOV outbreak variants only about a decade ago[2,48]. This finding suggests that EBOV Makona may be fairly new to West Africa, sharing recent common ancestry with Middle African variants that are found thousands of miles away. Molecular clock dating analyses have also shown that all recorded human EVD outbreaks caused by EBOV appear to share a common ancestor around 1975[2,51]. Notably, this is around the time of the first described EVD outbreak in 1976, suggesting that the EBOV lineage experienced a severe genetic bottleneck before the first human outbreak[52,53]. Despite their power[54], molecular clock dating studies of this type would undoubtedly benefit from additional EBOV genomic sequence data from both previous EVD outbreaks and animal reservoir populations.

Genetic diversification of Ebola virus Makona

Because of the relatively small magnitude and duration of previous EVD outbreaks, earlier EBOV sequencing efforts were necessarily limited to small numbers of temporally clustered cases. The data from these earlier studies largely comprised single viral lineages and led to the perception that EBOV genomes remain stable over the course of an outbreak[55–59]. However, the much larger size and duration of the 2013–2016 EVD epidemic (Fig. 1a, b) resulted in a different molecular epidemiological pattern for EBOV Makona, in which multiple virus lineages arose and co-circulated (Fig. 1c).

Despite their shared border, the EBOV Makona genomes sampled from the three most affected countries, Guinea, Sierra Leone and Liberia, generally (although not exclusively) form separate clusters on phylogenetic trees and exhibit different phylogenetic patterns[3,19–22,24–26] (Fig. 1b, c). Genomic studies have shown that the 2013–2016 EVD epidemic was dominated by three major lineages, denoted A (refs 21, 25), SL2 (ref. 2) and SL3 (refs 2, 3) (Fig. 1c). Most of these lineages—including lineage A[21,25] in Guinea, SL3 in Sierra Leone[3] and Liberian isolates[24]—circulated locally, with only sporadic cross-border transmission (Fig. 1c). By contrast, lineage SL2 (ref. 2) was the most widespread in the region[3,21,22,24,25] (Fig. 1c). This lineage probably arose in Sierra Leone[2] where it gave rise to lineage SL3 and several sub-lineages[3,22]. It crossed more than twice into Liberia[24], seeded several transmission chains in Guinea[21] and spread throughout Sierra Leone[2,3,22] (Fig. 1c). It is unclear whether any of these lineages carry mutations that could have affected their epidemic potential[60], or, perhaps more likely, whether the increased geographical spread of SL2 and SL3 is a reflection of chance epidemiological founding events (see below)[60,61].

Evolutionary dynamics of Ebola virus Makona

Although the origin and spread of the 2013–2016 EVD epidemic seem well resolved[1–3,20–22,24,25], other aspects of EBOV evolution during this epidemic have proven more controversial. A major point of contention in both scientific publications[2,3,20–22,24,62–64] and the popular press[65,66] has been whether the virus 'mutated' unusually rapidly during this outbreak. Unfortunately, much of this discussion is based on misrepresentations of what type of rate was measured and how these rates can be translated into predictions of phenotypic evolution.

The starting point for the debate over how quickly EBOV Makona evolved was the observation by Gire *et al.*[2] that the mean evolutionary rate early in the epidemic was $\sim 1.9 \times 10^{-3}$ (95% Bayesian credible interval: 1.11, 2.91×10^{-3}) nucleotide substitutions (subs) per site per year. This rate was approximately twice as high as that averaged from genomic sequences of EBOV variants sampled from multiple outbreaks, at around 0.9×10^{-3} (0.81, 1.18×10^{-3}) subs per site per year (ref. 2). EBOV outbreak variants are viral lineages responsible for human outbreaks. Other EBOV variants include EBOV Yambuku (Mayinga) from 1976, EBOV Kikwit from 1995, and EBOV Lomela from 2014 (DRC). The between-outbreak evolutionary rate therefore reflects estimates averaged across all EBOV variants. However, later studies of EBOV Makona consistently produced lower rate estimates than those generated by Gire *et al.*[3,21,22,25]. Indeed, taking the publicly available sequence data as a whole, estimates of the EBOV evolutionary rate for the 2013–2016 epidemic converge on a mean value of around 1.2×10^{-3} (1.13, 1.27×10^{-3}) (Fig. 2). The ensuing discussions of whether EBOV is evolving more or less rapidly than expected, and what this means for the ability of the virus to evolve changes in transmissibility and virulence, have become a common narrative[67–70].

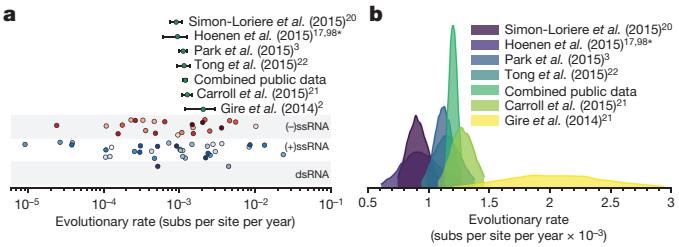


Figure 2 | Evolutionary rates of EBOV compared to those of other RNA viruses. a, Estimates of evolutionary rate in diverse RNA viruses. Green points at the top indicate the mean evolutionary rates estimated for EBOV during the 2013–2016 EVD epidemic from different studies, with solid lines showing the 95% credible intervals derived from BEAST analyses. Points at the bottom represent equivalent estimates (without uncertainty intervals) published previously for negative-sense single-strand RNA viruses (red), positive-sense single-strand RNA viruses (blue) and double-strand RNA viruses (purple)[72]. Points with the same shade belong to the same family. Evolutionary rate estimates for EBOV Makona occupy a narrow distribution within the range of rates observed in RNA viruses as a whole. **b**, 95% credible intervals for the distribution of evolutionary rates for EBOV from the 2013–2016 EVD epidemic published previously. *An erratum[98] revised the mean evolutionary rate estimate for ref. 17 (Hoenen *et al.*), to 1.32×10^{-3} (95% credible intervals: 0.89, 1.75×10^{-3}) subs per site per year.

The debate over the evolutionary dynamics of EBOV highlights a number of general issues in viral evolution. First, estimates of evolutionary rates are generally expected to be higher within outbreaks than between them. This is because the relatively short timescale over which sequences are sampled during outbreaks may be insufficient for mutations to be removed (or make them less likely to be fixed) by either natural selection or genetic drift. Hence, pathogen genomic sequences sampled early in epidemics will contain an excess of mildly deleterious variants that would eventually be eliminated by purifying selection[61]. This will tend to inflate evolutionary rates and in part explains why evolutionary rates in RNA viruses are often 'time-dependent': high towards the present, low towards the past[71,72]. Indeed, it is notable that as the 2013–2016 EVD epidemic progressed, analyses of evolutionary rate in EBOV converged on a reliable estimate (Fig. 2), which is expected owing to the increasing size of the data set combined with a longer sampling period. When viewed in the context of viruses as a whole, it is also striking that all the evolutionary rate estimates for EBOV fall in a narrow range towards the centre of a distribution that spans more than three orders of magnitude, from about 10^{-2} to about 10^{-5} subs per site per year (Fig. 2).

As well as time-dependence, it is possible that purifying selection may be relaxed in humans following cross-species transmission and/or that EBOV may undergo more replications per unit time during human outbreaks than in its reservoir species[2,73]. Both of these scenarios would increase the within-outbreak rate. Potential evidence for fundamental differences in evolutionary dynamics associated with species-jumping is provided by the EBOV Lomela variant that emerged in the DRC during 2014, causing a small EVD outbreak with 69 cases[74]. The branch length on the EBOV phylogenetic tree leading to the EBOV Lomela sequences from their common ancestor is far shorter than expected from their sampling time in 2014 (Fig. 3a), indicating a markedly lower evolutionary rate[74,75].

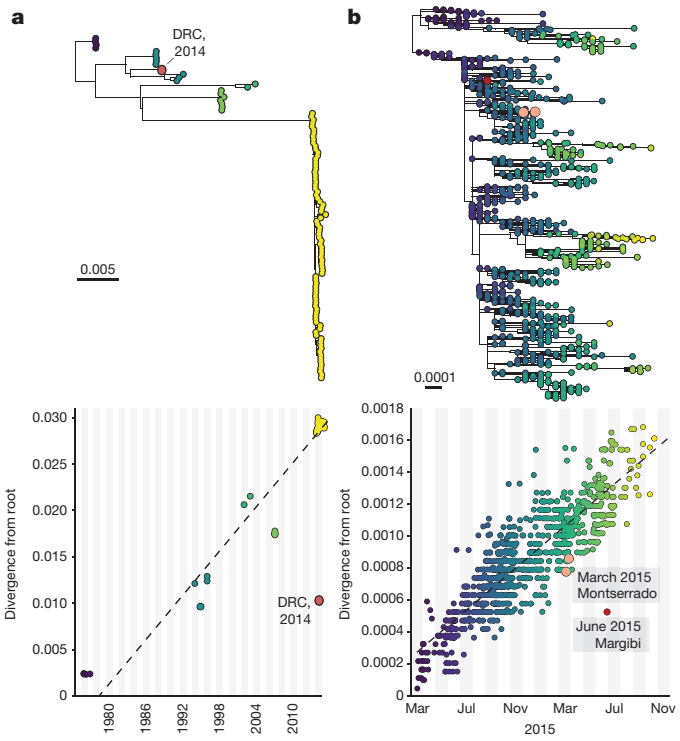


Figure 3 | Examples of violations of the Ebola virus molecular clock. **a**, Root-to-tip regression of genetic distances against time (month and year) of sampling for 105 representative EBOV variant sequences collected between 1976 and 2016 based on a maximum likelihood tree. **b** Equivalent root–to–tip regression of publicly available sequences from the 2013–2016 EVD epidemic using data on the day of sampling, and the maximum likelihood tree on which the estimates were made. RAxML (ref. 99) (panel **a**) and PhyML (ref. 100) (panel **b**) were used to estimate the maximum likelihood phylogenies under an HKY+Γ_4 substitution model that was rooted via least squares regression in TempEst. Substitutions accumulate linearly with time, with some variation. Sequences recovered from transmission events that occurred as a result of persistent EBOV infection often exhibit temporal anomalies. In this scenario, EBOV may accumulate substitutions at a lower rate during persistence in individuals compared to regular person-to-person transmission. Larger red points indicate sequences of EBOV sampled from EVD survivor-associated transmission chains[76,77]. Scale bar, nucleotide substitutions per site.

This could reflect an evolutionary history in a different reservoir host from those previously described for EBOV, in which replication rates and hence evolutionary rates are reduced, or in which purifying selection acts with greater potency than in humans. Lower EBOV evolutionary rates were also observed in suspected cases of transmission from human EVD survivors during the 2013–2016 epidemic (Fig. 3b)[76,77]. Unexpectedly low evolutionary rates may therefore serve as an important signal for detecting probable transmissions from EVD survivors during flare-ups[26,76] (Fig. 3b).

The debate over EBOV evolutionary rate estimates has also revealed confusion over the terminologies used to describe the rate at which genetic changes accumulate. The most straight-forward measure of the rate of molecular evolution is the nucleotide substitution rate. This rate describes the frequency with which mutations are fixed in populations through time and for EBOV is best approximated by the rate observed between outbreaks[61] (Box 1). This rate reflects the long-term evolutionary processes including selective constraints on the genome, host-species-specific adaptation and the cumulative results of genetic drift. In contrast, the rate of change within outbreaks might be better thought of as the evolutionary rate, as the short timescale of sampling necessarily means that not all mutations observed will be fixed. Both the substitution rate and evolutionary rate can be clearly distinguished from the mutation rate. This term relates to the rate at which mutations are generated during viral replication by intrinsic biochemical factors, and in particular to how frequently the viral polymerase makes errors[78] (Box 1, Box 1 Figure). This rate is generally challenging to measure[79,80] and is unknown for most viruses, including EBOV. It is therefore unfortunate that the debate over EBOV evolution has focused on 'mutation' (and hence potential differences intrinsic to particular virus lineages) when this is not the parameter that has been measured.

Finally, it is too simplistic to think that a twofold variation in rate estimates for EBOV will result in radically different evolutionary behaviour, especially when seen in the context of RNA viruses as a whole (Fig. 2). The likelihood of meaningful adaptive evolution depends not only on the rate at which the virus is able to generate mutations, but also on those environmental and host factors that shape the selection pressures acting on the virus. That filoviruses have infected a wide range of mammalian hosts[36,37,81] suggests that they are readily able to adapt to new environments irrespective of potential differences in evolutionary rate.

Phenotypic evolution of Ebola virus Makona

While the patterns of EBOV molecular evolution during the 2013–2016 EVD epidemic have been well characterized, it is not currently known whether any of the observed mutations have resulted in differences in viral phenotype. This is particularly the case with respect to such traits as antigenicity, transmissibility and virulence, or mutations that could have an impact on vaccines, therapeutics and diagnostics. Although genomic sequence data play a central role in understanding outbreak dynamics and evolution, revealing key aspects of viral phenotype using sequence data alone is fraught with difficulties, and may even be counterproductive to outbreak response by steering the focus away from more critical needs[65].

As the 2013–2016 West African epidemic of EVD was so much larger than previous outbreaks, it is possible that EBOV Makona possessed mutations that enhanced its transmissibility in humans. Without direct experimental data, however, a simpler scenario is that the scale and severity of the 2013–2016 EVD epidemic reflects a different epidemiological context than previous outbreaks. Under this model, most, if not all, EBOV variants entering human populations after cross-species transmission have the ability to cause major epidemics, but have been unable to do so because of a lack of a susceptible host population and/or environment. In particular, previous EVD outbreaks occurred in largely isolated and rural areas[41] (with the notable exception of the 1995 outbreak in Kikwit, which has a population of ~400,000; ref. 41), where there were either an insufficient number of susceptible people to guarantee long-term transmission, or the outbreak was quickly controlled by efficient interventions. The 2013–2016 EVD epidemic, in contrast, was the first in West Africa and the first in which a large EVD epidemic resulted in sustained community transmission from rural settings to major urban centres, where it was easier to establish large-scale transmission networks. This included the establishment of 'underground' networks, amplified by reluctance to seek medical advice in the affected communities, which greatly hindered intervention strategies focused on breaking chains of transmission. That the scale of the 2013–2016 EVD epidemic more reflects virus epidemiology rather than virus evolution is also supported by the failure to find evidence for heritable changes in the duration of virus shedding or virulence during the course of the 2013–2016 EVD epidemic[39,40,42–47].

However, it was also the case that EBOV evolution during the 2013–2016 EVD epidemic was characterized by an abundance of changes in the nucleotide and amino acid sequences that could fuel adaptation for more efficient human transmission; any mutations that increased R_0 would have been favoured by natural selection. Because of its key role in virus–host interactions, most attention has been directed towards the EBOV glycoprotein, and it is notable that the highest level of genetic amino acid diversity generated during the 2013–2016 EVD epidemic occurred in the glycoprotein (in particular, its mucin-like domain)[3,22]. For example, we observed 104 amino acid changes in glycoproteins that were shared by at least two EBOV Makona lineages from the 1,500 available EBOV genomes that make up 5%

BOX 1

Different measures of genome sequence change

Mutation rate

As viruses replicate, mutational errors are incorporated into the viral genome. The mutation rate is therefore typically expressed as the number of mutations per site, per replication event. The mutation rate for RNA viruses such as EBOV is largely determined by the viral RNA-dependent RNA polymerase, which lacks proof-reading activity. The estimation of mutation rates requires complex sequencing-based or phenotypic marker experiments that correct for the impact of natural selection[79]. Mutation rates are unknown for most viruses and have not been determined for EBOV, although it would be predicted to be comparable to other (−)ssRNA viruses and is likely to be similar across all EBOV outbreak variants.

Evolutionary rate

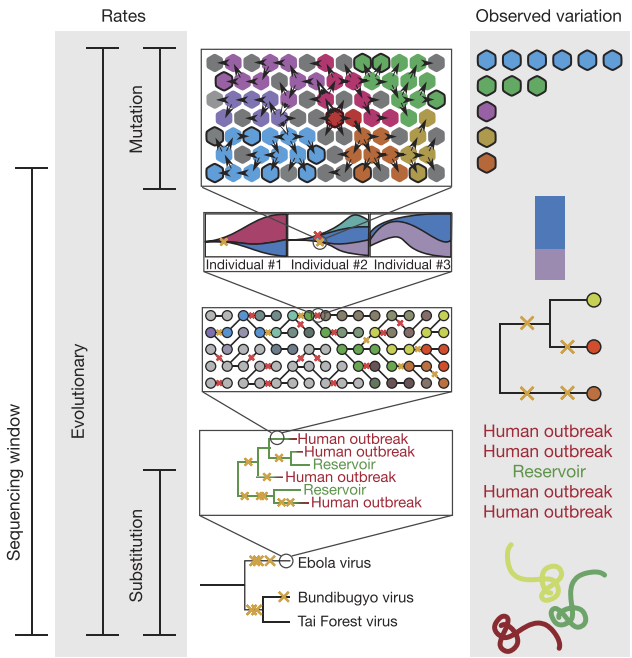
The evolutionary rate of a virus can be defined as the observed rate at which new variants arise and spread in the viral population. This can be measured by methods that compare the genetic change in viral genomes collected at different times. Importantly, evolutionary rates in RNA viruses may be dependent on the timescale over which they are measured: they are elevated in the short-term, such as within disease outbreaks, because mildly deleterious mutations may not have been eliminated by purifying selection[71,72].

Substitution rate

The substitution rate is best described as the long-term rate at which genetic variants become fixed in a virus lineage over evolutionary time-scales, such as between human outbreaks in the case of EBOV. Hence, this rate reflects the complex interplay of natural selection, genetic drift, modes of transmission and epidemiological processes. This rate will also usually be lower than the short-term evolutionary rate because many of the variants circulating within outbreaks and epidemics will ultimately be eliminated. Furthermore, saturation—repeated changes at the same site—will further reduce the measured substitution rate.

Fixation rate

An added complexity in estimating rates in RNA viruses is that the population genetic concept of 'fixation', central to the definition of substitution, is ill-defined. In slowly evolving organisms, fixation events can usually be distinguished from polymorphisms by analysing individual nucleotide sites within and between species. However, in rapidly evolving RNA viruses, fixation can be described to occur (1) at the level of individual hosts over the course of infection, (2) in viral lineages within specific geographic locations or epidemiological networks (such as the different lineages of EBOV generated during the 2013–2016 EVD epidemic), (3) in global meta-populations, and (4) between different viral species.



Box 1 Figure | Illustration of different measures of genomic variation. Mutations accumulate over time. This phenomenon is at the core of molecular clocks, a class of methods that aim to convert molecular phylogenies with branch lengths given in expected substitutions per site into plausible temporal phylogenies in which branch lengths are given in time units and the trees themselves are embedded in time. By making use of sequences sampled at different times, such methods can estimate the evolutionary rate that provides the conversion from genetic distance into time. As phylogenetic methods have become ever more powerful and easily accessible, confusion has resulted from the frequent and interchangeable use of the terms mutation rate and substitution rate to signify the 'molecular clock' rate. Mutation and substitution rates, however, sit on the opposite ends of the evolutionary rate continuum and neither is the appropriate term for the molecular clock rate derived from densely sequenced epidemics.

of the more than 28,000 reported EVD cases[7]. While it is not known whether any of these amino acid changes led to functional differences, one plausibly important glycoprotein variant that originated early in the epidemic (in lineage SL2)[24] is an alanine to valine change at residue 82 (A82V). This is the first substitution observed in the receptor binding domain of EBOV and could potentially alter the interaction between the EBOV glycoprotein and its host receptor Niemann-Pick C1 (NPC1)[82,83]. Clearly, determining whether lineages of EBOV Makona carrying A82V or other mutations that arose during the 2013–2016 EVD epidemic differ in epidemic potential should be a research priority.

Irrespective of potential differences in transmissibility that are yet to be uncovered, it is more certain that EBOV Makona is no different from previous EBOV outbreak variants when it comes to bodily fluids being the primary route of transmission[81]. Early on in the 2013–2016 EVD epidemic there was high-profile speculation that EBOV could evolve

respiratory (that is, airborne) transmission due to genetic diversity in the viral population[62,64,65]. However, there is no evidence for airborne EBOV transmission during the 2013–2016 EVD epidemic—or any other EVD outbreaks—and nor are there any examples of other viruses evolving a new mode of transmission on the timescale of individual outbreaks. Although influenza virus shifts its mode of transmission from (primarily) faecal–oral in its wild bird reservoir to respiratory in humans[84], this change occurs at the point of cross-species transmission and not during human outbreaks.

While the occurrence of airborne transmission can be eliminated for EBOV, studies using genomic sequence data have conclusively shown that sexual transmission plays a previously unappreciated role for EBOV dissemination and reignition[77,85–88]. However, the long-term epidemiological and evolutionary implications of this mode of transmission are unclear and warrant further in-depth studies.

Public health implications of genomic epidemiology

In addition to providing essential information on the pattern and dynamics of viral evolution during epidemics, viral genomic data may be of more direct public health importance. Indeed, the 2013–2016 EVD epidemic is arguably the first in which genomic data have been used directly in a real-time public health setting, to inform policies and infection control[2,7,25,26]. That some of these studies were undertaken under difficult field conditions[19,23,25,26] highlights the potential for portable genomic sequencing to transform outbreak responses[7,25].

The simplest use of genomic data during outbreaks has been to reveal the pathways of viral spread through communities; when combined with phylogeographic approaches[22,89], the results can be used to direct intervention methods to transmission hot-spots and to determine the impact of specific interventions such as border closures. For example, Tong and colleagues used viral genome sequencing to show how EBOV spread from the capital city of Freetown to multiple districts throughout Sierra Leone[22], with Arias *et al.* later documenting how virus traffic from Freetown established new transmission clusters late in the epidemic[26]. Similarly, phylogenetic analyses revealed the co-circulation of multiple EBOV lineages within individual localities such as Conakry[20], as well as cross-border virus traffic between Guinea and Sierra Leone[25], highlighting important gaps in intervention. On a more localized epidemiological scale, genome sequence data provide a way to reveal who infected whom in EBOV transmission networks (although see below). Pathogen sequence data can therefore yield key information on the likelihood of, for example, sexual transmission[77,85–88], as well as the possible transmission of EBOV via breast milk[26]. A similarly precise reconstruction of transmission chains is essential in understanding the multiple reignition events that occurred during the EBOV epidemic and their relation to viral transmissions from EVD survivors[76] (Fig. 3b). It is unclear whether the small subset of EVD survivors that harbour persistent infections pose a sustained infection risk or whether an episode of renewed viral replication is required for transmission to occur. Considering the pattern and degree of EBOV genetic change within such cases may provide critical insights. Phylogenetic approaches also provide a powerful way to accurately estimate various outbreak parameters, such as R_0, including that for individual virus lineages that are slow and difficult to obtain using longitudinal case data[90,91]. Finally, virus 'super spreaders' within human populations can also be readily identified using pathogen sequence data[92].

Despite the quantity and quality of the viral genome sequence data generated during the 2013–2016 EVD epidemic, there are limitations to the scope and impact of genomic epidemiology. Clearly, the direct phenotypic effects of individual mutations on vaccines, therapeutics and diagnostics need to be tested experimentally. However, should viral lineages that differ in such properties arise during outbreaks, evolutionary genomic analyses can provide a powerful means to both determine their origins and rapidly track their spread through human populations.

Lessons learned and future directions

The 2013–2016 EVD epidemic has set the benchmark for the use of large-scale molecular epidemiology as an essential tool in outbreak response. Given the development of portable sequencing technologies, real-time viral genome sequencing is now possible in clinics and diagnostic laboratories, including in resource-limited settings[25,93]. This will offer critical data to inform epidemiological intervention, but will require a willingness to invest in scientific infrastructure, healthcare and training of local staff in the affected countries[94]. The need for immediate analysis and the growth of open sharing of sequence data means the challenge in genomic studies may be moving from data acquisition to analysis and interpretation. However, it is also the case that in-country real-time sequencing was not established until relatively late in the West African epidemic[19,23,25,26,93], when case numbers had already begun to decline. In addition, many of the genome sequences were obtained in the absence of strong clinical and epidemiological metadata, such as the precise geographical location from where the sample was obtained, whether the individual survived the infection, and the time to the onset of symptoms. While it may be difficult to obtain such data during a rapidly developing outbreak, this limits the usefulness of genomic sequencing data in addressing a number of central biological questions, such as the virological basis to any variation in disease presentation and the evolution of pathogen virulence. An important lesson for the study and management of future disease outbreaks is not only that portable sequencing platforms should be deployed as rapidly as possible, but that each sequence obtained should be linked to as much relevant metadata as is ethically and technically possible.

Despite the insights provided by the analysis of EBOV genome data, it is also clear that major questions remain. For future outbreaks it will be important to resolve exact chains of transmission (that is, who infected whom), as this provides vital information on the patterns and mechanisms of virus spread within single communities and hospitals, which will help target interventions. Sequence data from the 2013–2016 EVD epidemic indicated that these chains were difficult to infer using the population consensus sequences from individual hosts, although in several cases they were shown to be in agreement with epidemiological studies[2,3,25]. Hence, although EBOV evolves rapidly, mutations are not necessarily fixed at the scale of individual transmission events, which limits phylogenetic resolution. One solution is to examine the transmission patterns of intra-host single nucleotide variants (iSNVs) (ref. 3). If multiple iSNVs are routinely transmitted between individuals (that is, that EBOV is not subject to a severe population bottleneck at inter-host transmission) then tracking the inheritance patterns of these variants can provide information on how transmission patterns exist between individual hosts, as previously shown for influenza virus[95,96]. Importantly, it has been shown that substantial intra-host variation can be observed for EBOV, with 2–5 iSNVs per infected patient being typical when using a minor allele frequency cutoff of 5%[2–4,26]. As the cutoff is lowered, the numbers of observed iSNVs increase sharply[2,3].

Genomic studies undertaken in West Africa towards the end of the 2013–2016 EVD outbreak illustrated how iSNV data can help to resolve EBOV transmission pathways. For example, Arias and colleagues showed how the analysis of iSNVs from EBOV patients in Sierra Leone could provide strong support for sexual transmission from EVD survivors[26]. Determining the number of iSNVs that transmit between hosts can also provide key information on the severity of the transmission bottleneck[3], itself critical for understanding the ability of natural selection to shape patterns of genetic diversity.

Finally, while large-scale EBOV sequence studies have now been undertaken in human populations, there is an evident and critical need to determine the ecology and evolution of EBOV in its animal reservoir(s). While most current data points to bats being the ultimate reservoir host[34,35], long-term studies of EBOV in bats have yet to be performed and it is likely that other host species exist, which may have a major bearing on epidemiological dynamics. To truly understand the ecology and evolution of EBOV, as well as its mechanisms of pathogenicity, will require information on the virus in all its host–virus interactions, and not just those associated with EVD outbreaks in humans.

Received 15 January; accepted 23 August 2016.

1. Baize, S. *et al.* Emergence of Zaire Ebola virus disease in Guinea. *N. Engl. J. Med.* **371,** 1418–1425 (2014).
   The first paper to describe the emergence of EBOV Makona in Guinea in December 2013, providing the sequences of three full-length viral genomes.
2. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345,** 1369–1372 (2014).
   Obtained the first large-scale, near-real-time EBOV genomic data from 78 patients in Sierra Leone, which provided critical insights into virus spread during the early stages of the epidemic.
3. Park, D. J. *et al.* Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell* **161,** 1516–1526 (2015).
4. Andersen, K. G. *et al.* Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell* **162,** 738–750 (2015).
5. Kuhn, J. H. *et al.* 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 (2014).
6. WHO. Statement on the 1st meeting of the IHR Emergency Committee on the 2014 Ebola outbreak in West Africa. *WHO Media Centre* http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/(2014).
7. Ebola Situation Report, W. H. O. *Ebola virus disease outbreak* http://apps.who.int/ebola/current-situation/ebola-situation-report-30-march-2016 (2016).
8. Nichol, S. T. *et al.* Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* **262,** 914–917 (1993).
9. Holmes, E. C. *et al.* The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. *J. Infect. Dis.* **171,** 45–53 (1995).
10. Tsui, S. K., Chim, S. S. & Lo, Y. M. Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. *N. Engl. J. Med.* **349,** 187–188 (2003).
11. Ghedin, E. *et al.* Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437,** 1162–1166 (2005).
12. Garten, R. J. *et al.* Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* **325,** 197–201 (2009).
13. Chiu, C. Y. Viral pathogen discovery. *Curr. Opin. Microbiol.* **16,** 468–478 (2013).
14. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).
15. Metzker, M. L. Sequencing technologies—the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).
16. Mardis, E. R. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* **6,** 287–303 (2013).
17. Hoenen, T. *et al.* Virology. Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science* **348,** 117–119 (2015).
18. Bell, A. *et al.* Genome sequence analysis of Ebola virus in clinical samples from three British healthcare workers, August 2014 to March 2015. *Euro Surveill.* **20,** 21131 (2015).
19. Kugelman, J. R. *et al.* Monitoring of Ebola virus Makona evolution through establishment of advanced genomic capability in Liberia. *Emerg. Infect. Dis.* **21,** 1135–1143 (2015).
20. Simon-Loriere, E. *et al.* Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature* **524,** 102–104 (2015).
21. Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature* **524,** 97–101 (2015).
    The first study to obtain a large EBOV sequence data set collected over multiple outbreaks, providing insights into the phylogenetic relationship between EBOV outbreak variants as well as other filoviruses.
22. Tong, Y. G. *et al.* Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature* **524,** 93–96 (2015).
23. Smits, S. L. *et al.* Genotypic anomaly in Ebola virus strains circulating in Magazine Wharf area, Freetown, Sierra Leone, 2015. *Euro Surveill.* **20,** 30035 (2015).
24. Ladner, J. T. *et al.* Evolution and spread of Ebola virus in Liberia, 2014–2015. *Cell Host Microbe* **18,** 659–669 (2015).
25. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530,** 228–232 (2016).
    Implemented real-time sequencing in Guinea using the Oxford Nanopore MinION technology, which enabled EBOV genomic sequence data to be obtained within hours or days of cases being detected.
26. Arias, A. *et al.* Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol.* **2,** vew016(2016).
    The largest single study of EBOV genomics to date. Implemented real-time sequencing in Sierra Leone using the Thermo Fisher Ion Torrent platform, which enabled EBOV genomic sequence data to be obtained within days of cases being detected.
27. Matranga, C. B. *et al.* Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.* **15,** 519 (2014).
28. Woolhouse, M. E., Rambaut, A. & Kellam, P. Lessons from Ebola: improving infectious disease surveillance to inform outbreak management. *Sci. Transl. Med.* **7,** 307rv5 (2015).
29. Neher, R. A. & Bedford, T. Nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* **31,** 3546–3548 (2015).
30. Bedford, T. & Neher, R. A. NextStrain—real-time analysis of Ebola virus evolution. *NextStrain* http://ebola.nextstrain.org/ (2015).
31. Lewis, J. A. *et al.* Phylogenetic relationships of dengue-2 viruses. *Virology* **197,** 216–224 (1993).
32. Ray, S. C., Arthur, R. R., Carella, A., Bukh, J. & Thomas, D. L. Genetic epidemiology of hepatitis C virus throughout Egypt. *J. Infect. Dis.* **182,** 698–707 (2000).
33. WHO. WHO experts consultation on Ebola Reston pathogenicity in humans. *Emergencies Preparedness, Response* http://www.who.int/csr/resources/publications/WHO_HSE_EPR_2009_2/en/ (2015).
34. Leroy, E. M. *et al.* Fruit bats as reservoirs of Ebola virus. *Nature* **438,** 575–576 (2005).
    Obtained sequence information from three species of fruit bats collected during the 2001–2003 EVD outbreak in Gabon to provide the first convincing evidence implicating fruit bats as a/the likely EBOV reservoir.
35. Leendertz, S. A., Gogarten, J. F., Düx, A., Calvignac-Spencer, S. & Leendertz, F. H. Assessing the evidence supporting fruit bats as the primary reservoirs for Ebola viruses. *EcoHealth* **13,** 18–25 (2016).
36. Taylor, D. J., Leach, R. W. & Bruenn, J. Filoviruses are ancient and integrated into mammalian genomes. *BMC Evol. Biol.* **10,** 193 (2010).
37. Taylor, D. J., Ballinger, M. J., Zhan, J. J., Hanzly, L. E. & Bruenn, J. A. Evidence that ebolaviruses and cuevaviruses have been diverging from marburgviruses since the Miocene. *PeerJ* **2,** e556 (2014).
38. Report of an International Commission. Ebola haemorrhagic fever in Zaire, 1976. *Bull. World Health Organ.* **56,** 271–293 (1978).
39. WHO Ebola Response Team. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371,** 1481–1495 (2014).
40. Kucharski, A. J. & Edmunds, W. J. Case fatality rate for Ebola virus disease in West Africa. *Lancet* **384,** 1260 (2014).
41. CDC. Outbreaks chronology: Ebola virus disease. *Outbreaks* http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html (2015).
42. Marzi, A. *et al.* Delayed disease progression in Cynomolgus macaques infected with Ebola virus Makona strain. *Emerg. Infect. Dis.* **21,** 1777–1783 (2015).
43. Schieffelin, J. S. *et al.* Clinical illness and outcomes in patients with Ebola in Sierra Leone. *N. Engl. J. Med.* **371,** 2092–2100 (2014).
44. Volz, E. & Pond, S. Phylodynamic analysis of Ebola virus in the 2014 Sierra Leone epidemic. *PLoS Curr.* **6,** ecurrents.outbreaks.6f7025f1271821d4c815385b08f5f80e (2014).
45. Alizon, S., Lion, S., Murall, C. L. & Abbate, J. L. Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. *Virulence* **5,** 825–827 (2014).
46. Althaus, C. L. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Curr.* **6,** ecurrents.outbreaks.91afb5e0f279e7f29e7056095255b288 (2014).
47. Rosello, A. *et al.* Ebola virus disease in the Democratic Republic of the Congo, 1976-2014. *eLife* **4,** e09015 (2015).
48. Dudas, G. & Rambaut, A. Phylogenetic analysis of Guinea 2014 EBOV ebolavirus outbreak. *PLoS Curr.* **6,** ecurrents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d (2014).
49. Leroy, E. M. *et al.* Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science* **303,** 387–390 (2004).
50. Wittmann, T. J. *et al.* Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. *Proc. Natl Acad. Sci. USA* **104,** 17123–17127 (2007).
51. Walsh, P. D., Biek, R. & Real, L. A. Wave-like spread of Ebola Zaire. *PLoS Biol.* **3,** e371 (2005).
52. Carroll, S. A. *et al.* Molecular evolution of viruses of the family *Filoviridae* based on 97 whole-genome sequences. *J. Virol.* **87,** 2608–2616 (2013).
53. Biek, R., Walsh, P. D., Leroy, E. M. & Real, L. A. Recent common ancestry of Ebola Zaire virus found in a bat reservoir. *PLoS Pathog.* **2,** e90 (2006).
    Using previously available sequence data, this paper describes a strong link between EBOV found in bats and EBOV found in humans during outbreaks, suggesting common ancestry within the last three decades.
54. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4,** e88 (2006).
55. Sanchez, A., Trappier, S. G., Mahy, B. W., Peters, C. J. & Nichol, S. T. 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 (1996).
56. Georges-Courbot, M. C. *et al.* Isolation and phylogenetic characterization of Ebola viruses causing different outbreaks in Gabon. *Emerg. Infect. Dis.* **3,** 59–62 (1997).
57. Georges, A. J. *et al.* Ebola hemorrhagic fever outbreaks in Gabon, 1994–1997: epidemiologic and health control issues. *J. Infect. Dis.* **179** (Suppl. 1), S65–S75 (1999).
58. Rodriguez, L. L. *et al.* Persistence and genetic stability of Ebola virus during the outbreak in Kikwit, Democratic Republic of the Congo, 1995. *J. Infect. Dis.* **179** (Suppl. 1), S170–S176 (1999).
59. Grard, G. *et al.* Emergence of divergent Zaire ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. *J. Infect. Dis.* **204** (Suppl. 3), S776–S784 (2011).
60. Łuksza, M., Bedford, T. & Lässig, M. Epidemiological and evolutionary analysis of the 2014 Ebola virus outbreak. Preprint at: http://arxiv.org/pdf/1411.1722.pdf (2014).
61. Holmes, E. C. *The Evolution and Emergence of RNA Viruses.* (Oxford Univ. Press, 2009).

62. Osterholm, M. T. *et al.* Transmission of Ebola viruses: what we know and what we do not know. *MBio* **6,** e00137 (2015).
63. Ponce De Leon-Rosales, S., Arredondo-Hernandez, R., Macias, A. & Wenzel, R. P. Ebola, through air or not through air: that is the question. *Front. Public Health* **2,** 292 (2015).
64. Leroy, E. M., Labouba, I., Maganga, G. D. & Berthet, N. Ebola in West Africa: the outbreak able to change many things. *Clin. Microbiol. Infect.* **20,** 0597–0599 (2014).
65. Osterholm, M. T. What we're afraid to say about Ebola. *NY Times* (11 Sept. 2014).
66. Basch, C. H., Basch, C. E. & Redlener, I. Coverage of the Ebola virus disease epidemic in three widely circulated United States newspapers: implications for preparedness and prevention. *Health Promot. Perspect.* **4,** 247–251 (2014).
67. Callaway, E. M. Ebola's fast evolution questioned. *NATNEWS* doi:10.1038/nature.2015.17200 (2015).
68. Vogel, G. A reassuring snapshot of Ebola. *Science* **347,** 1407 (2015).
69. Saphire, E. O. New advances in the effort against Ebola. *Cell Host Microbe* **17,** 545–547 (2015).
70. Liu, S. Q., Rayner, S. & Zhang, B. How Ebola has been evolving in West Africa. *Trends Microbiol.* **23,** 387–388 (2015).
71. Ho, S. Y., Shapiro, B., Phillips, M. J., Cooper, A. & Drummond, A. J. Evidence for time dependency of molecular rate estimates. *Syst. Biol.* **56,** 515–522 (2007).
72. Duchêne, S., Holmes, E. C. & Ho, S. Y. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc.* **281,** 20140732(2014).
73. Moya, A., Holmes, E. C. & González-Candelas, F. The population genetics and evolutionary epidemiology of RNA viruses. *Nat. Rev. Microbiol.* **2,** 279–288 (2004).
74. Maganga, G. D. *et al.* Ebola virus disease in the Democratic Republic of Congo. *N. Engl. J. Med.* **371,** 2083–2091 (2014).
75. Lam, T. T., Zhu, H., Chong, Y. L., Holmes, E. C. & Guan, Y. Puzzling origins of the Ebola outbreak in the Democratic Republic of the Congo, 2014. *J. Virol.* **89,** 10130–10132 (2015).
76. Blackley, D. J. *et al.* Reduced evolutionary rate in reemerged Ebola virus transmission chains. *Sci. Adv.* **2,** e1600378 (2016).
Using EBOV sequencing from seven EVD cases collected during a 'flare-up' in Liberia, this study investigates how these cases were linked to individuals who had survived EVD and carried the virus as a persistent asymptomatic infection.
77. Mate, S. E. *et al.* Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* **373,** 2448–2454 (2015).
Using a combination of epidemiological and sequence-based investigations, this study convincingly shows that EBOV is capable of sexual transmission many months after an EVD survivor was discharged from the hospital.
78. Malpica, J. M. *et al.* The rate and character of spontaneous mutation in an RNA virus. *Genetics* **162,** 1505–1511 (2002).
79. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84,** 9733–9748 (2010).
80. Domingo-Calap, P. & Sanjuán, R. Experimental evolution of RNA versus DNA viruses. *Evolution* **65,** 2987–2994 (2011).
81. Kuhn, J. H. Filoviruses. A compendium of 40 years of epidemiological, clinical, and laboratory studies. *Arch. Virol. Suppl.* **20,** 13–360 (2008).
82. Côté, M. *et al.* Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature* **477,** 344–348 (2011).
83. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477,** 340–343 (2011).
84. Horimoto, T. & Kawaoka, Y. Influenza: lessons from past pandemics, warnings from current incidents. *Nat. Rev. Microbiol.* **3,** 591–600 (2005).
85. Deen, G. F. *et al.* Ebola RNA persistence in semen of Ebola virus disease survivors—preliminary report. *N. Engl. J. Med.* (2015). 10.1056/NEJMoa1511410
86. Christie, A. *et al.* Possible sexual transmission of Ebola virus—Liberia, 2015. MMWR Morb. Mortal. *Wkly. Rep.* **64,** 479–481 (2015).
87. Fischer, R. J., Judson, S., Miazgowicz, K., Bushmaker, T. & Munster, V. J. Ebola virus persistence in semen *ex vivo*. *Emerg. Infect. Dis.* **22,** 289–291 (2016).
88. Thorson, A., Formenty, P., Lofthouse, C. & Broutet, N. Systematic review of the literature on viral persistence and sexual transmission from recovered Ebola survivors: evidence and recommendations. *BMJ Open* **6,** e008859 (2016).
89. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5,** e1000520 (2009).
90. Stadler, T. *et al.* Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* **29,** 347–357 (2012).
91. Stadler, T., Kühnert, D., Rasmussen, D. A. & du Plessis, L. Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. *PLoS Curr.* **6,** ecurrents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f (2014).
Using sequence data from 72 EVD cases from Sierra Leone, this study utilized phylodynamic analyses to estimate several epidemiological parameters, obtaining an R_0 of around 2.18.
92. Stadler, T. & Bonhoeffer, S. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Phil. Trans. R. Soc. Lond. B* **368,** 20120198 (2013).
93. Hoenen, T. *et al.* Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerg. Infect. Dis.* **22,** 331–334 (2016).
94. Yozwiak, N. L. *et al.* Roots, not parachutes: research collaborations combat outbreaks. *Cell* **166,** 5–8 (2016).
95. Poon, L. L. *et al.* Quantifying influenza virus diversity and transmission in humans. *Nat. Genet.* **48,** 195–200 (2016).
96. Stack, J. C., Murcia, P. R., Grenfell, B. T., Wood, J. L. & Holmes, E. C. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. *Proc. R. Soc.* **280,** 20122173 (2013).
97. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
98. Hoenen, T. *et al.* Erratum: Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science* **348,** aac5674 (2015).
99. Stamatakis, A., Ludwig, T. & Meier, H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21,** 456–463 (2005).
100. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).

Acknowledgements We thank all the African doctors, nurses, scientists, and outbreak responders who worked to control the 2013–2016 EVD epidemic, some of whom tragically died in the process. We also thank the EBOV genome sequence data producers for making their data publicly available, S. Schaffner for suggestions and reading of the manuscript, and L. M. Carvalho for donating evolutionary rate data. E.C.H. is funded by an NHMRC Australia Fellowship (AF30). G.D. is supported by EU (FP7/2007-2013) Grant Agreement no. 278433-PREDEMICS and the Mahan Postdoctoral Fellowship from the Computational Biology Program at Fred Hutchinson Cancer Research Center. A.R. is supported by EU (FP7/2007-2013) Grant Agreement no. 278433-PREDEMICS, H2020 Grant Agreement no. 643476-COMPARE, and a Wellcome Trust Strategic Award (VIZIONS; 093724). K.G.A. is a PEW Biomedical Scholar, and his work is supported by an NIH National Center for Advancing Translational Studies Clinical and Translational Science Award UL1TR001114, and NIAID contract HHSN272201400048C.

Author Contributions All authors were involved in data analysis and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.C.H. (edward.holmes@sydney.edu.au) or K.G.A. (kristian@andersen-lab.com).

Reviewer Information *Nature* thanks C. Drosten, P. Lemey, G. Palacios and T. Sadler for their contribution to the peer review of this work.