The Evolution of Ebola virus: Insights from the 2013–2016 Epidemic

Edward C. Holmes[1,*], **Gytis Dudas**[2,3], **Andrew Rambaut**[3,4,5], and **Kristian G. Andersen**[6,7,8,9,*]

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Biological Sciences and Sydney Medical School, Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia

[2]Fred Hutchinson Cancer Research Center, Seattle, WA 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, MD 20892, USA

[6]The Scripps Research Institute, Department of Immunology and Microbial Science, La Jolla, CA 92037, USA

[7]The Scripps Research Institute, Department of Integrative Structural and Computational Biology, La Jolla, CA 92037, USA

[8]Scripps Translational Science Institute, La Jolla, CA 92037, USA

[9]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Preface

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. Herein, 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.

Introduction

The 2013–2016 Ebola virus disease (EVD) epidemic in West Africa appears to have begun following human contact with an animal (likely bat) reservoir of Ebola virus (EBOV) in December 2013, in the small village of Meliandou in Guéckédou Prefecture, Guinea[1]. After

*To whom correspondence should be addressed: ECH: edward.holmes@sydney.edu.au; KGA: kristian@andersen-lab.com.

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 make 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 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. In the following we illustrate 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.

Review

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 only been confirmed in 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].

Ebola virus disease 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] (although CFRs are difficult to estimate for EVD[39,40], so should be interpreted with caution). Between 1977 and 2014, 12 smaller outbreaks were reported in Middle Africa, with 32 to 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 where 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 ~70%[1,39,40,43] and estimates for the basic reproduction number (R_0) fall between 1.5–2.5, 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. Due to the high genetic similarity of virus genomes sampled from the beginning of the epidemic, a single spill-over infection seems the most likely[1,2]. Phylogenetic analyses also make it clear that once the outbreak was established, later lineages of EBOV Makona have descended from those circulating earlier in the epidemic (Fig. 1c)[2,3,20–22,24,25]. 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 place the first case around late December 2013 in Guinea[1]. In particular, 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 suggests that EBOV Makona may be fairly new to West Africa, sharing recent common ancestry with Middle African variants that are 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]. Strikingly, this is around the time of the first described EVD outbreak in 1976, suggesting that the EBOV lineage

experienced a severe genetic bottleneck prior to 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 during the 2013–2016 epidemic

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 (Fig. 1b, c)[3,19–22,24–26]. Genomic studies have shown that the 2013–2016 EVD epidemic was dominated by three major lineages, denoted 'A'[21,25], 'SL2'[2], and 'SL3'[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 transmissions (Fig. 1c). In contrast, lineage SL2[2] has been the most widespread in the region (Fig. 1c)[3,21,22,24,25]. This lineage likely 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 later section below)[60,61].

Evolutionary dynamics of Ebola virus during the 2013–2016 epidemic

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 evolves was the observation by Gire *et al.* that the mean evolutionary rate early in the epidemic was ~1.9×10^{-3} [95% Bayesian credible interval: 1.11, 2.91×10^{-3}] nucleotide substitutions per site, per year (subs/site/year)[2]. This rate was approximately twice as high as that averaged from genomic sequences of EBOV variants[a] sampled from multiple outbreaks, at ~0.9×10^{-3} [0.81, 1.18×10^{-3}] subs/site/year[2]. However, later studies of EBOV Makona consistently produced lower rate estimates than that 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 outbreak converge on a mean value of ~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].

The debate over the evolutionary dynamics of EBOV highlights a number of general issues in viral evolution. First, estimates of evolutionary rate are generally expected to be higher within than between outbreaks. This is because the relatively short time-scale over which sequences are sampled during outbreaks may be insufficient for mutations to be removed (or less likely be fixed) by either natural selection or genetic drift. Hence, pathogen genomic sequences sampled early within 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 toward the present, low toward 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 due 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 center of a distribution that spans more than three orders of magnitude, from ~10^{-2} to ~10^{-5} subs/site/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 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]. This could reflect an evolutionary history in a different reservoir host to those previously described for EBOV, in which replication rates and hence evolutionary rates are reduced, or where purifying selection acts with greater potency than in humans. Intriguingly, 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 of 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

---

[a]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.

rate', which describes the frequency that mutations are fixed in populations through time and for EBOV is best approximated by the rate observed between outbreaks (Box 1)[61]. This rate reflects the long-term evolutionary process of selective constraints on the genome, host species-specific adaptation and the cumulative results of genetic drift. The rate of change within outbreaks might be better thought of as the 'evolutionary rate' as the short time-scale 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, particularly how frequently the viral polymerase makes errors[78] (Box 1). 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 two-fold variation in rate estimates for EBOV will result in radically different evolutionary behavior, especially when seen in the context of RNA viruses as a whole (Fig. 2). The likelihood of meaningful adaptive evolution not only depends 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 during the 2013–2016 epidemic

While the patterns of EBOV molecular evolution during the 2013–2016 EVD epidemic have been well characterized, it is currently unknown whether any of the observed mutations have resulted in differences in viral phenotype – particularly with respect to such traits as antigenicity, transmissibility, and virulence – or 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 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 possesses mutations that have 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 following cross-species transmissions have the ability to cause major epidemics, but have been unable to do so because of a lack of 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[41]), where there were either an insufficient number of susceptibles 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 centers, where it was easier to establish large-scale transmission networks. This included the establishment of 'underground' networks, amplified by reluctance in the affected communities, that 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 is also the case that EBOV evolution during the 2013–2016 EVD epidemic is characterized by an abundance of nucleotide and amino acid sequence that could fuel adaptation for more efficient human transmission; any mutations that increased R_0 would have been favored by natural selection. Because of its key role in virus-host interactions, most attention has been directed toward the EBOV glycoprotein (GP), and it is notable that the highest level of genetic amino acid diversity generated during the 2013–2016 EVD epidemic occurred in the GP (in particular, its mucin-like domain)[3,22]. For example, we observed 104 amino acid changes in GP that are shared by at least two EBOV Makona lineages across the ~1,500 publicly available complete EBOV genomes that make up about 5% of the more than 28,000 reported EVD cases[7]. While it is not known whether any of these amino acid changes lead to functional differences, one plausibly important GP 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 GP and its host receptor 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 yet to be uncovered, it is more certain that EBOV Makona is no different than 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 (*i.e.*, 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 – nor are there any examples of other viruses evolving a new mode of transmission on the time-scale of individual outbreaks. Although influenza virus shifts its mode of transmission from (primarily) fecal-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 utilizing 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, helping 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 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 transmissions[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 (Fig. 3b)[76]. It is unclear whether the small subset of EVD survivors that harbor 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, the identification of virus 'super spreaders' within human populations can also be readily achieved based on 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, and therapeutics, diagnostics need to be tested experimentally. However, should viral lineages that differ in such properties arise during outbreaks, evolutionary genomic analyses 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 viable in clinics and diagnostic laboratories, including in resource-limited settings[25,93]. This will offer critical information 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 (*i.e.*, 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)[3]. If multiple iSNVs are routinely transmitted between individuals (*i.e.*, 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 significant intra-host variation can be observed for EBOV, with 2–5 iSNVs per infected patient 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 toward the end of the 2013–2016 EVD outbreak illustrated how iSNV data can help resolve EBOV transmission pathways. For example, Arias and colleagues showed how the analysis of iSNVs from EBOV patients in Sierra Leone can 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.

Acknowledgments

References

1. Baize S, et al. Emergence of Zaire Ebola virus disease in Guinea. N Engl J Med. 2014; 371:1418–1425. The first paper to describe the emergence of EBOV Makona in Guinea in December 2013, providing the sequences of three full-length viral genomes. [PubMed: 24738640]

2. Gire SK, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014; 345:1369–1372. 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. [PubMed: 25214632]

3. Park DJ, et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. Cell. 2015; 161:1516–1526. [PubMed: 26091036]

4. Andersen KG, et al. Clinical sequencing uncovers origins and evolution of Lassa virus. Cell. 2015; 162:738–750. [PubMed: 26276630]

5. Kuhn JH, 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. 2014; 6:4760–4799. [PubMed: 25421896]

6. WHO. Statement on the 1st meeting of the IHR Emergency Committee on the 2014 Ebola outbreak in West Africa. WHO Media Centre. 2014. Available at: http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/

7. WHO. Ebola Situation Report – 30 March 2016. Ebola virus disease outbreak. 2016. Available at: http://apps.who.int/ebola/current-situation/ebola-situation-report-30-march-2016

8. Nichol ST, et al. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. Science. 1993; 262:914–917. [PubMed: 8235615]

9. Holmes EC, et al. The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. J Infect Dis. 1995; 171:45–53. [PubMed: 7798682]

10. Tsui SK, Chim SS, Lo YM, Chinese, University of Hong Kong Molecular SARS Research Group. Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. N Engl J Med. 2003; 349:187–188. [PubMed: 12853594]

11. Ghedin E, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature. 2005; 437:1162–1166. [PubMed: 16208317]

12. Garten RJ, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science. 2009; 325:197–201. [PubMed: 19465683]

13. Chiu CY. Viral pathogen discovery. Curr Opin Microbiol. 2013; 16:468–478. [PubMed: 23725672]

14. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008; 26:1135–1145. [PubMed: 18846087]

15. Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet. 2010; 11:31–46. [PubMed: 19997069]

16. Mardis ER. Next-generation sequencing platforms. Annu Rev Anal Chem. 2013; 6:287–303.

17. Hoenen T, et al. Virology. Mutation rate and genotype variation of Ebola virus from Mali case sequences. Science. 2015; 348:117–119. [PubMed: 25814067]

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. 2015; 20:21131. ii. [PubMed: 26027482]

19. Kugelman JR, et al. Monitoring of Ebola virus Makona evolution through establishment of advanced genomic capability in Liberia. Emerg Infect Dis. 2015; 21:1135–1143. [PubMed: 26079255]

20. Simon-Loriere E, et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. Nature. 2015; 524:102–104. [PubMed: 26106863]

21. Carroll MW, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature. 2015; 524:97–101. 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. [PubMed: 26083749]

22. Tong YG, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature. 2015; 524:93–96. [PubMed: 25970247]

23. Smits SL, et al. Genotypic anomaly in Ebola virus strains circulating in Magazine Wharf area, Freetown, Sierra Leone, 2015. Euro Surveill. 2015; 20

24. Ladner JT, et al. Evolution and spread of Ebola virus in Liberia, 2014–2015. Cell Host Microbe. 2015; 18:659–669. [PubMed: 26651942]

25. Quick J, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016; 530:228–232. 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. [PubMed: 26840485]

26. Arias A, et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Virus Evolution. 2016; 2:vew016. 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. [PubMed: 28694998]

27. Matranga CB, et al. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol. 2014; 15:519. [PubMed: 25403361]

28. Woolhouse ME, Rambaut A, Kellam P. Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management. Sci Transl Med. 2015; 7:307rv5.

29. Neher RA, Bedford T. Nextflu: Real-time tracking of seasonal influenza virus evolution in humans. Bioinformatics. 2015; 31:3546–3548. [PubMed: 26115986]

30. Bedford, T., Neher, RA. NextStrain – real-time analysis of Ebola virus evolution. NextStrain. 2015. Available at: http://ebola.nextstrain.org/

31. Lewis JA, et al. Phylogenetic relationships of dengue-2 viruses. Virology. 1993; 197:216–224. [PubMed: 8212556]

32. Ray SC, Arthur RR, Carella A, Bukh J, Thomas DL. Genetic epidemiology of hepatitis C virus throughout egypt. J Infect Dis. 2000; 182:698–707. [PubMed: 10950762]

33. WHO. WHO experts consultation on Ebola Reston pathogenicity in humans. Emergencies preparedness, response. 2015. Available at: http://www.who.int/csr/resources/publications/WHO_HSE_EPR_2009_2/en/

34. Leroy EM, et al. Fruit bats as reservoirs of Ebola virus. Nature. 2005; 438:575–576. 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. [PubMed: 16319873]

35. Leendertz SA, Gogarten JF, Düx A, Calvignac-Spencer S, Leendertz FH. Assessing the evidence supporting fruit bats as the primary reservoirs for Ebola viruses. Ecohealth. 2015; 13:18–25. [PubMed: 26268210]

36. Taylor DJ, Leach RW, Bruenn J. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol. 2010; 10:193. [PubMed: 20569424]

37. Taylor DJ, Ballinger MJ, Zhan JJ, Hanzly LE, Bruenn JA. Evidence that ebolaviruses and cuevaviruses have been diverging from marburgviruses since the Miocene. PeerJ. 2014; 2:e556. [PubMed: 25237605]

38. Commission, Report of an International. Ebola haemorrhagic fever in Zaire, 1976. Bull World Health Organ. 1978; 56:271–293. [PubMed: 307456]

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. 2014; 371:1481–1495. [PubMed: 25244186]

40. Kucharski AJ, Edmunds WJ. Case fatality rate for Ebola virus disease in West Africa. Lancet. 2014; 384:1260.

41. CDC. Outbreaks chronology: Ebola virus disease. Outbreaks. 2015. Available at: http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html

42. Marzi A, et al. Delayed disease progression in Cynomolgus Macaques infected with Ebola virus Makona strain. Emerg Infect Dis. 2015; 21:1777–1783. [PubMed: 26402165]

43. Schieffelin JS, et al. Clinical illness and outcomes in patients with Ebola in Sierra Leone. N Engl J Med. 2014; 371:2092–2100. [PubMed: 25353969]

44. Volz E, Pond S. Phylodynamic analysis of Ebola virus in the 2014 Sierra Leone epidemic. PLoS Curr. 2014; 6 ecurrents.outbreaks.6f7025f1271821d4c815385b08f5f80e.

45. Alizon S, Lion S, Murall CL, Abbate JL. Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. Virulence. 2014; 5:825–827. [PubMed: 25495064]

46. Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS Curr. 2014; 6 ecurrents.outbreaks.91afb5e0f279e7f29e7056095255b288.

47. Rosello A, et al. Ebola virus disease in the Democratic Republic of the Congo, 1976–2014. Elife. 2015; 4:e09015. [PubMed: 26525597]

48. Dudas G, Rambaut A. Phylogenetic analysis of Guinea 2014 EBOV ebolavirus outbreak. PLoS Curr. 2014; 6 ecurrents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

49. Leroy EM, et al. Multiple Ebola virus transmission events and rapid decline of central African wildlife. Science. 2004; 303:387–390. [PubMed: 14726594]

50. Wittmann TJ, et al. Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. Proc Natl Acad Sci U S A. 2007; 104:17123–17127. [PubMed: 17942693]

51. Walsh PD, Biek R, Real LA. Wave-like spread of Ebola Zaire. PLoS Biol. 2005; 3:e371. [PubMed: 16231972]

52. Carroll SA, et al. Molecular evolution of viruses of the family Filoviridae based on 97 whole-genome sequences. J Virol. 2013; 87:2608–2616. [PubMed: 23255795]

53. Biek R, Walsh PD, Leroy EM, Real LA. Recent common ancestry of Ebola Zaire virus found in a bat reservoir. PLoS Pathog. 2006; 2:e90. Using previously available sequence data, this paper describes a strong link between EBOV found in bats and EBOV from in humans during outbreaks, suggesting common ancestry within the last three decades. [PubMed: 17069458]

54. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006; 4:e88. [PubMed: 16683862]

55. Sanchez A, Trappier SG, Mahy BW, Peters CJ, Nichol ST. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. Proc Natl Acad Sci U S A. 1996; 93:3602–3607. [PubMed: 8622982]

56. Georges-Courbot MC, et al. Isolation and phylogenetic characterization of Ebola viruses causing different outbreaks in Gabon. Emerg Infect Dis. 1997; 3:59–62. [PubMed: 9126445]

57. Georges AJ, et al. Ebola hemorrhagic fever outbreaks in Gabon, 1994–1997: epidemiologic and health control issues. J Infect Dis. 1999; 179(Suppl 1):S65–75. [PubMed: 9988167]

58. Rodriguez LL, et al. Persistence and genetic stability of Ebola virus during the outbreak in Kikwit, Democratic Republic of the Congo, 1995. J Infect Dis. 1999; 179(Suppl 1):S170–6. [PubMed: 9988181]

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. 2011; 204(Suppl 3):S776–84. [PubMed: 21987750]

60. Łuksza M, Bedford T, Lässig M. Epidemiological and evolutionary analysis of the 2014 Ebola virus outbreak. arXiv [q-bio.PE]. 2014

61. Holmes, EC. The Evolution and Emergence of RNA Viruses. Oxford University Press; 2009.

62. Osterholm MT, et al. Transmission of Ebola viruses: what we know and what we do not know. MBio. 2015; 6:e00137. [PubMed: 25698835]

63. Ponce De Leon-Rosales S, Arredondo-Hernandez R, Macias A, Wenzel RP. Ebola, through air or not through air: That is the question. Front Public Health. 2014; 2:292. [PubMed: 25646157]

64. Leroy EM, Labouba I, Maganga GD, Berthet N. Ebola in West Africa: the outbreak able to change many things. Clin Microbiol Infect. 2014; 20:O597–9. [PubMed: 25204860]

65. Osterholm MT. What we're afraid to say about Ebola. NY Times. 2014 The Opinion Pages.

66. Basch CH, Basch CE, Redlener I. Coverage of the ebola virus disease epidemic in three widely circulated United States newspapers: implications for preparedness and prevention. Health Promot Perspect. 2014; 4:247–251. [PubMed: 25649411]

67. Callaway EM. Ebola's fast evolution questioned. Nature News. 2015

68. Vogel G. A reassuring snapshot of Ebola. Sci News. 2015

69. Saphire EO. New advances in the effort against Ebola. Cell Host Microbe. 2015; 17:545–547. [PubMed: 25974296]

70. Liu SQ, Rayner S, Zhang B. How Ebola has been evolving in West Africa. Trends Microbiol. 2015; 23:387–388. [PubMed: 25986056]

71. Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. Syst Biol. 2007; 56:515–522. [PubMed: 17562475]

72. Duchêne S, Holmes EC, Ho SY. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc Biol Sci. 2014; 281

73. Moya A, Holmes EC, González-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol. 2004; 2:279–288. [PubMed: 15031727]

74. Maganga GD, et al. Ebola virus disease in the Democratic Republic of Congo. N Engl J Med. 2014; 371:2083–2091. [PubMed: 25317743]

75. Lam TT, Zhu H, Chong YL, Holmes EC, Guan Y. Puzzling origins of the Ebola outbreak in the Democratic Republic of the Congo, 2014. J Virol. 2015; 89:10130–10132. [PubMed: 26202242]

76. Blackley DJ, et al. Reduced evolutionary rate in reemerged Ebola virus transmission chains. Science Advances. 2016; 2:e1600378. 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. [PubMed: 27386513]

77. Mate SE, et al. Molecular evidence of sexual transmission of Ebola virus. N Engl J Med. 2015; 373:2448–2454. 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. [PubMed: 26465384]

78. Malpica JM, et al. The rate and character of spontaneous mutation in an RNA virus. Genetics. 2002; 162:1505–1511. [PubMed: 12524327]

79. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010; 84:9733–9748. [PubMed: 20660197]

80. Domingo-Calap P, Sanjuan R. Experimental evolution of RNA versus DNA viruses. Evolution. 2011; 65:2987–2994. [PubMed: 21967437]

81. Kuhn JH. Filoviruses. A compendium of 40 years of epidemiological, clinical, and laboratory studies. Arch Virol Suppl. 2008; 20:13–360. [PubMed: 18637412]

82. Côté M, et al. Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. Nature. 2011; 477:344–348. [PubMed: 21866101]

83. Carette JE, et al. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature. 2011; 477:340–343. [PubMed: 21866103]

84. Horimoto T, Kawaoka Y. Influenza: Lessons from past pandemics, warnings from current incidents. Nat Rev Microbiol. 2005; 3:591–600. [PubMed: 16064053]

85. Deen GF, et al. Ebola RNA persistence in semen of Ebola virus disease survivors – preliminary report. N Engl J Med. 2015; doi: 10.1056/NEJMoa1511410

86. Christie A, et al. Possible sexual transmission of Ebola virus – Liberia, 2015. MMWR Morb Mortal Wkly Rep. 2015; 64:479–481. [PubMed: 25950255]

87. Fischer RJ, Judson S, Miazgowicz K, Bushmaker T, Munster VJ. Ebola virus persistence in semen ex vivo. Emerg Infect Dis. 2016; 22:289–291. [PubMed: 26811984]

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. 2016; 6:e008859.

89. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009; 5:e1000520. [PubMed: 19779555]

90. Stadler T, et al. Estimating the basic reproductive number from viral sequence data. Mol Biol Evol. 2012; 29:347–357. [PubMed: 21890480]

91. Stadler T, Kühnert D, Rasmussen DA, du Plessis L. Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. PLoS Curr. 2014; 6 ecurrents.outbreaks. 02bc6d927ecee7bbd33532ec8ba6a25f. 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. Philos Trans R Soc Lond B Biol Sci. 2013; 368:20120198. [PubMed: 23382421]

93. Hoenen T, et al. Nanopore sequencing as a rapidly deployable Ebola outbreak tool. Emerg Infect Dis. 2016; 22

94. Yozwiak NL, et al. Roots, Not Parachutes: Research Collaborations Combat Outbreaks. Cell. 2016; 166:5–8. [PubMed: 27368093]

95. Poon LL, et al. Quantifying influenza virus diversity and transmission in humans. Nat Genet. 2016; 48:195–200. [PubMed: 26727660]

96. Stack JC, Murcia PR, Grenfell BT, Wood JL, Holmes EC. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. Proc Biol Sci. 2013; 280:20122173. [PubMed: 23135678]

97. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29:1969–1973. [PubMed: 22367748]

98. Hoenen, T., et al. Erratum for the report 'Mutation rate and genotype variation of Ebola virus from Mali case sequences'. In: Hoenen, T.Safronetz, D.Groseth, A.Wollenberg, KR.Koita, OA.Diarra, B.Fall, IS.Haidara, FC.Diallo, F.Sanogo, M.Sarro, YS.Kone, A.Togo, ACG.Traore, A.Kodio, M.Dosseh, A.Rosenke, K.de Wit, E.Feldmann, F.Ebihara, H.Munster, VJ.Zoon, KC.Feldmann, H., Sow, S., editors. Science. Vol. 348. 2015. p. aac5674

99. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 2005; 21:456–463. [PubMed: 15608047]

100. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]
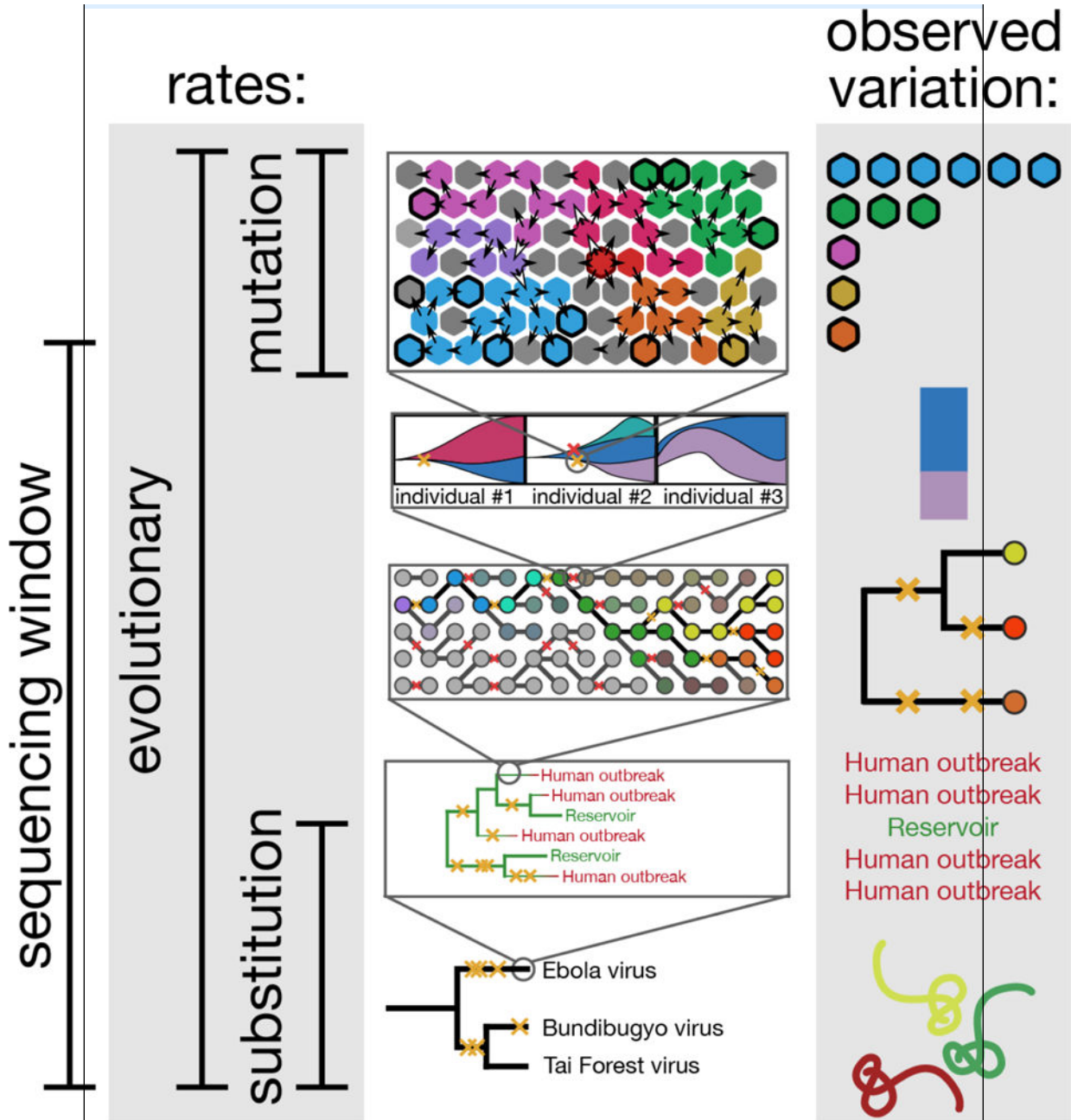
Box 1

Figure. Different measures of genome sequence change

Mutations accumulate with 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 where 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.

| Concept	Explanation	"holmes_figure_box.pdf"
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 likely 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 as the result of a complex interplay of natural selection, genetic drift, modes of transmission and epidemiological processes. 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 time-scale 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. This rate will 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.	
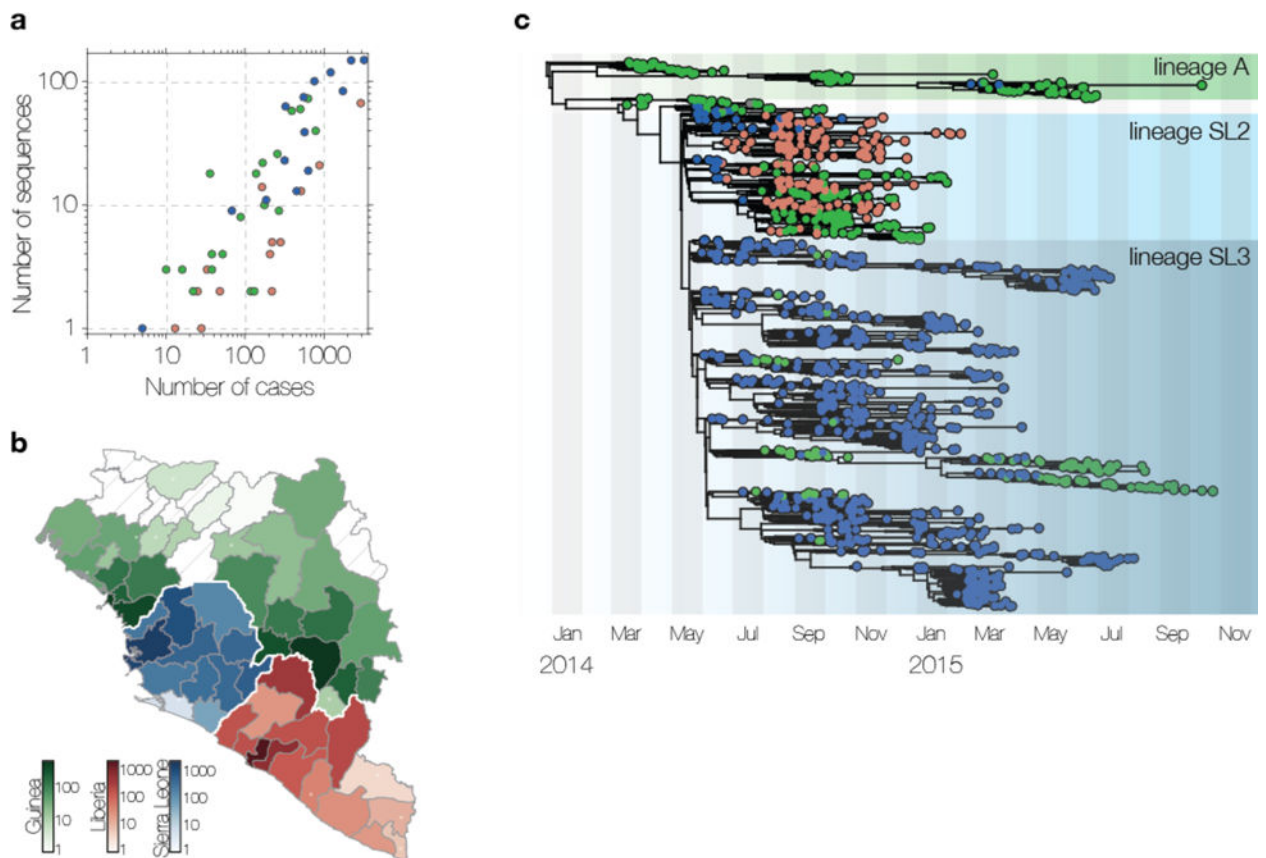
Figure 1. Evolution of EBOV during the 2013–2106 outbreak showing the extent and location of virus sampling

(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 colored 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 colored backgrounds. The sequence alignment was partitioned into 4 categories: codon positions 1, 2 and 3, and non-coding intergenic regions. Changes in each of the 4 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 colored according to the country where the patient was most likely infected: green for Guinea, red for Liberia and blue for Sierra Leone. All panels are current as of April 19th, 2016.
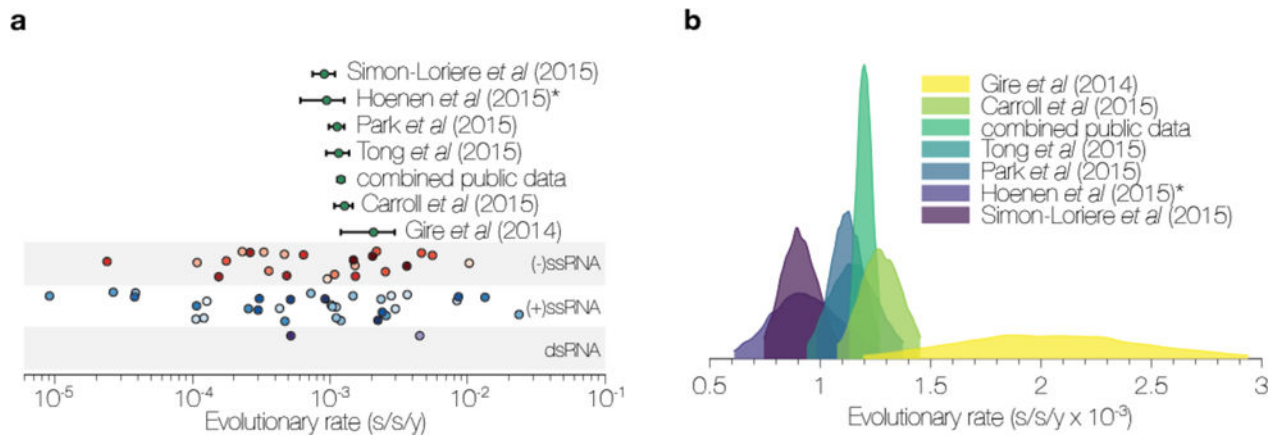
Figure 2. Evolutionary rates of EBOV compared to those of other RNA viruses
(a) Estimates of evolutionary rate in diverse RNA viruses. Green points at the top indicate
the mean evolutionary rates estimated for EBOV during the 2013–2016 EVD epidemic from
different studies, with dashed lines showing the 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 colored
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] to 1.32×10^{-3} [95% credible intervals: 0.89, 1.75 $\times$
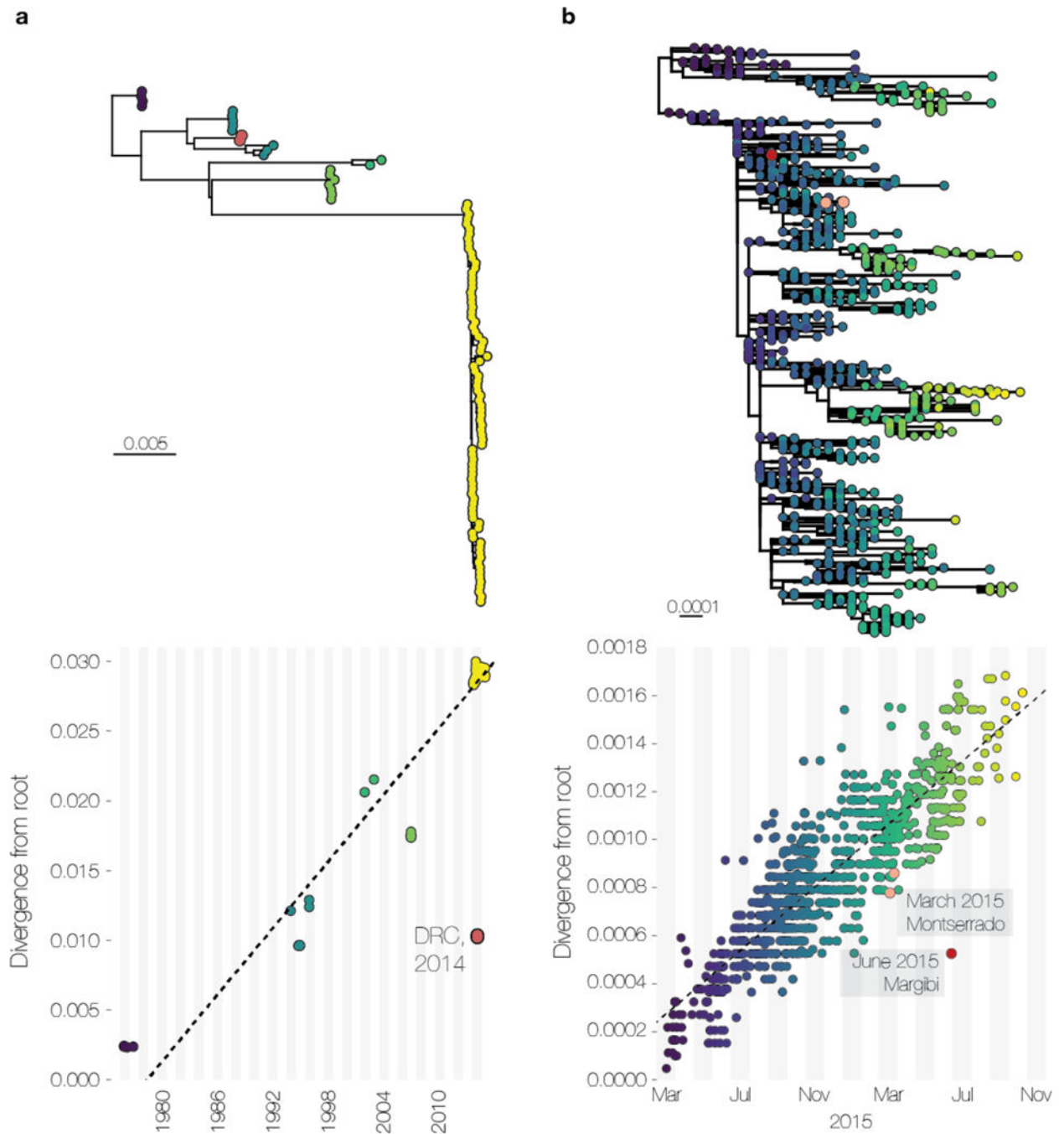10^{-3}] subs/site/year.

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 utilizing data on the day of sampling, and the maximum likelihood tree on which the estimates were made. RAxML[99] (panel A) and PhyML[100] (panel B) were used to find the maximum likelihood phylogenies under an HKY $+\Gamma_4$ substitution model which 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].

Table 1

Overview over EBOV sequencing studies performed during the 2013–2016 epidemic.

Study	Platform	Method	Sequencing location	Case location	# Seqs
Baize S et al. N Engl J Med. 2014 Apr	Sanger	Amplified	International	Guinea	3
Gire SK et al. Science. 2014 Sep	Illumina	Direct	International	Sierra Leone	79
Hoenen T et al. Science. 2015 Apr	Sanger	Amplified	International	Mali	4
Bell A et al. Euro Surveill. 2015 May	Illumina	Direct	International	UK	3
Park DJ et al. Cell. 2015 Jun	Illumina	Direct	International	Sierra Leone	232
Kugelman JR et al. Emerg Infect Dis. 2015 Jul	Illumina	Direct	In-country/Liberia	Liberia	25
Simon-Loriere E et al. Nature. 2015 Aug	Illumina	Direct	International	Guinea	85
Carroll MW et al. Nature. 2015 Aug	Illumina	Direct	International	Guinea/Liberia	179
Tong YG et al. Nature. 2015 Aug	BGISEQ-100	Amplified	?	Sierra Leone	175
Smits et al. Euro Surveil. 2015 Sep	Ion Torrent	Amplified	In-country/Sierra Leone	Sierra Leone	49
Ladner JT et al. Cell Host Microbe. 2015 Dec	Illumina	Direct	International	Liberia	140
Quick J et al. Nature. 2016 Feb	MinION	Amplified	In-country/Guinea	Guinea	137
Hoenen T et al. Emerg Infect Dis. 2016 Feb	MinION	Amplified	In-country/Liberia	Liberia	8
Arias et al Virus Evolution 2016 Jun	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/enrichment; Amplified = material amplified via amplicon-based PCR prior to viral sequencing), sequencing location, country of origin for the sequenced samples (case location), and number of EBOV genomes produced.