Evidence of latency reshapes our understanding of Ebola virus reservoir dynamics

John T. McCrone[1] Guy Baele[2] Ifeanyi F. Omah[3]
Eddy Kinganda-Lusamaki[4,5] Joseph A. Brew[6] Luiz M. Carvalho[7]
Gytis Dudas[8] Placide Mbala-Kingebeni[4,9] Marc A. Suchard[10,11,12]
Andrew Rambaut[3]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA
[2]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium
[3]Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK
[4]Institut National de Recherché Biomedicale, University of Kinshasa, Kinshasa, Democratic Republic of the Congo
[5]TransVIHMI, Université de Montpellier, Montpellier, France
[6] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[7]School of Applied Mathematics, Getulio Vargas Foundation (FGV), Rio de Janeiro, Brazil
[8]Institute of Biotechnology, Life Sciences Centre, Vilnius University, Vilnius, Lithuania
[9]South African National Bioinformatics Institute, University of the Western Cape, South Africa
[10]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA
[11]Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
[12]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

October 17, 2025

Abstract

Ebola virus (EBOV) has caused severe outbreaks of haemorrhagic fever in Central and West Africa since the first observed zoonotic epidemic in the late 1970s. While recent outbreaks have revealed much about the epidemiological dynamics that sustain human-to-human transmission, the mechanisms by which the virus persists between outbreaks are unknown. Previously, phylogenetic approaches have been used to characterise the EBOV reservoir from the evolutionary relationships among observed human outbreaks. We here employ a novel phylogenetic latency model – inspired by recent observations of extreme EBOV evolutionary rate heterogeneity in humans – to characterise the natural history of EBOV and by extension its reservoir. We find the prevailing model of EBOV reservoir dynamics is deficient, and the long-term EBOV evolutionary rate is slower than previously believed. The hypothesis that EBOV diversity dates back to a bottleneck event just prior to the first human outbreak is not supported by the data. Further, our results suggest that EBOV undergoes extended periods of quiescence in the reservoir, similar to that observed in a small fraction of human infections. These findings have significant implications for understanding the source of EBOV outbreaks, characterising the EBOV reservoir, and uncovering the factors that contribute to EBOV outbreaks in humans.

Introduction

Ebola virus (EBOV) is a filovirus of the genus *Ebola* which causes Ebola virus disease (EVD), a severe haemorrhagic fever in humans. The first recorded outbreak of EBOV occurred in 1976 near Yambuku, in the north of what is now the Democratic Republic of the Congo (DRC) (*1*). The

epidemic spread through the Yambuku Mission Hospital resulting in 318 cases of EVD and 280 deaths. Because much of the transmission was nosocomial and the disease was unknown to the people in the region, epidemiologists suspected the virus was imported into the clinic by a traveller from Sudan, where a separate outbreak of haemorrhagic fever had also been reported. However, the WHO bulletin report from the outbreak noted that EBOV may be endemic to the area and reside in an unknown reservoir due to the discovery of anti-EBOV glycoprotein (GP) antibodies in five individuals who were not ill and did not have a connection with the recorded outbreak (*1*). Later, it was shown that the Sudan outbreak was caused by a separate filovirus – Ebola virus Sudan – further supporting a separate zoonotic origin of the outbreak in Yambuku (*2*). Since EBOV was first isolated in 1976, there have been at least 18 detected outbreaks sparked by independent zoonoses. It is now widely accepted that EBOV is endemic in the wildlife of Central Africa. However, the exact reservoir of EBOV remains uncharacterised.

EBOV is thought to be maintained among bat populations in Central Africa, based largely on observations that Marburg virus, a related filovirus, transmits among fruit bats (*3*, *4*). Field work supports this hypothesis. Small fragments ($\sim$265bp) of EBOV genomes as well as anti-Ebola GP antibodies were isolated from fruit bats sampled near human outbreaks along the border of the Republic of the Congo and Gabon in the early 2000s (*5*). Additionally, in 2019 EcoHealthAlliance sequenced a fragment of the virus from a greater long-fingered bat (*Miniopterus inflatus*) in Liberia (*6–8*).

But the reservoir is likely more complex than one species or population. Partial sequences isolated from gorilla, chimpanzee, and duiker carcasses associated with human outbreaks in Central Africa, and serology from non-human primates suggest multiple species act as intermediate hosts (*9–11*). The degree to which any particular species contributes to the long-term reservoir is unknown. Neither live virus, nor whole genome has been isolated from a bat population, and concentrated sampling efforts near recent outbreaks have failed to find an active reservoir (*12*). Much of what is known about the EBOV reservoir has been derived from evolutionary analyses, which impute population dynamics from the ancestral relationships among zoonotic spill-overs (*11*, *13*, *14*). In the absence of any EBOV genome sequences from infected reservoir species, we are observing the evolutionary history of EBOV in non-human animals entirely from outbreaks and zoonotic infections in humans.

It has long been suggested that EBOV diversity dates back to the 1970s outbreaks of Yambuku (1976) and nearby Bonduni (1977) (*13–17*). This prevailing view relies heavily on a temporal rooting of the EBOV phylogeny estimated from isolates sampled between the 1976 outbreak and the West African outbreak of 2013. Phylogeographic models based on this origin found EBOV exhibited wave-like spread out of the northern DRC, and proposed an active expanding epidemic within the unobserved reservoir (*14*). The location of the 2013 West African outbreak drew this geographic model into question, and led some to propose a new root position based on outgroup rooting with other *Ebola* species. This analysis found the West African outbreak as an outgroup to all other EBOV lineages (*18*). However, outgroup rooting with divergent species is unreliable (*19*), and while the epidemic in West Africa was a geographic outlier (Supplemental Figure S1A), its divergence from the 1976 Yambuku outbreak is consistent with previous estimates of the virus's evolutionary rate (Figure 1A) (*20*). The prevailing model, based on a temporal rooting, has persisted to this day.

There have been eight sequenced zoonotic outbreaks of EBOV since 2013 (*21–25*) (Supplemental Table S1). The most recent occurred in August, 2025 (*26*). All have been less diverged than expected given the accepted model (Figure 1), leading some to hypothesize that EBOV may have entered a new reservoir population with slower replication dynamics (*21*). However, recent observations of EBOV rate heterogeneity in humans offer a different perspective. There are now multiple, well-documented cases in which EBOV has persisted for extended periods of time (weeks to years) in a recovered individual, before sparking a secondary transmission chain (*27–30*). The maximum length of EBOV persistence in humans is unknown. The most extreme case to date lasted six years (*30*), suggesting EBOV may persist in hosts for much longer than initially thought. Genomic se-

quencing has revealed that fewer mutations accumulate during these long periods of persistence than expected given the EBOV evolutionary rate (*28*, *30*). How infectious EBOV persists in a host for long periods of time without mutating (and by extension replicating) is unknown. The epidemiology of these well-documented events implies it must.

Observations of the persistence phenomenon in survivors of acute EVD suggest that, within those individuals, persistence is a two-state process with replication (and thus mutation) only occurring during the initial infection and at the point of transmission (either before or after). Recrudescence of EVD — the re-emergence of replicating virus — after a period of remission has been observed (*31*), but does not seem to be required as most secondary transmission chains have begun through sexual transmission without re-establishment of replication in the survivor. For the Guinea 2021 outbreak, linked to the 2013-2016 epidemic by a period of over six years, the index case suffered acute, and fatal, EVD but was not known to have been infected previously nor to have had contacts who were survivors (*30*). In this case, the virus exhibited almost no genetic divergence from the isolates taken during the epidemic suggesting most of the 6 year period was spent in a non-replicating, latent, state.

We hypothesize that the decreased divergence associated with recent zoonotic events results from a latent process in the EBOV reservoir similar to that observed in persistent human infections, and that accounting for these dynamics could reconcile recent observations with the long-held model of EBOV's evolutionary past. Here, we have developed a model of the virus's evolutionary rate that accounts for periods of latency, and applied it to a dataset representing EBOV evolution in the reservoir. Our findings suggest EBOV latency is not unique to human infections, but rather has occurred throughout the natural history of EBOV. Furthermore, accounting for these dynamics draws the prevailing model of long-term EBOV evolution into question and redefines our expectations for the timing and location of future outbreaks.

Results

We collected a representative set of Ebola virus genomes, taken from 19 plausibly independent zoonotic spill-over events (Table S2). The genomes included in this study were sequenced from human subjects, but because we restrict our sampling to one genome per outbreak and prioritize early samples, our phylogeny is largely shaped by the evolutionary dynamics in the reservoir.

Restricting our dataset to genomes from the eleven recorded zoonotic outbreaks between 1976 and 2013, and rooting near the 1976 Yambuku outbreak recovers the accepted model of EBOV evolution and yields a root-to-tip regression with a high correlation coefficient and an R^2 of 0.92 (Figure 1A & C). This rooting also produces an evolutionary rate of 7.15×10^{-4} substitutions per site per year (s/s/y), not dissimilar to the 1.3×10^{-3} s/s/y observed during human outbreaks (*32*).

However, as Lam *et al.* (*21*) and later Mbala-Kingebeni *et al.* (*33*) observed, this visually appealing linear relationship breaks down with the 2014 DRC and subsequent outbreaks (Figure 1C). Attempts by us (*34*) and others (*33*) to account for the limited divergence of recent outbreaks have required topological constraints and/or informative priors on the molecular clock rate in addition to *a priori* assigned local clock models in order to recover a root near the original 1976 Yambuku outbreak. The necessity of these informative priors in addition to *a priori* assigned local clocks suggests the temporal signal in the first 11 outbreaks is not as strong as initially thought.

Evidence for an alternative root position

It has been noted elsewhere (*33*) that the EBOV phylogeny is made up of several clades of outbreaks that cluster by time and location (denoted by colour in Figure 1 and S1). We observed that the root-to-tip regression based on a root near the 1976 outbreak is driven by the long branches that connect these clusters (Figure 1A and C).

However, there seems to be an alternative regression within each cluster (coloured clades in Figure 1), that differs consistently from this trend. The rate implied by these independent regression slopes is slower than that expected if EBOV diversity descends from a common ancestor near the 1976 Yambuku outbreak (Figure 1C).

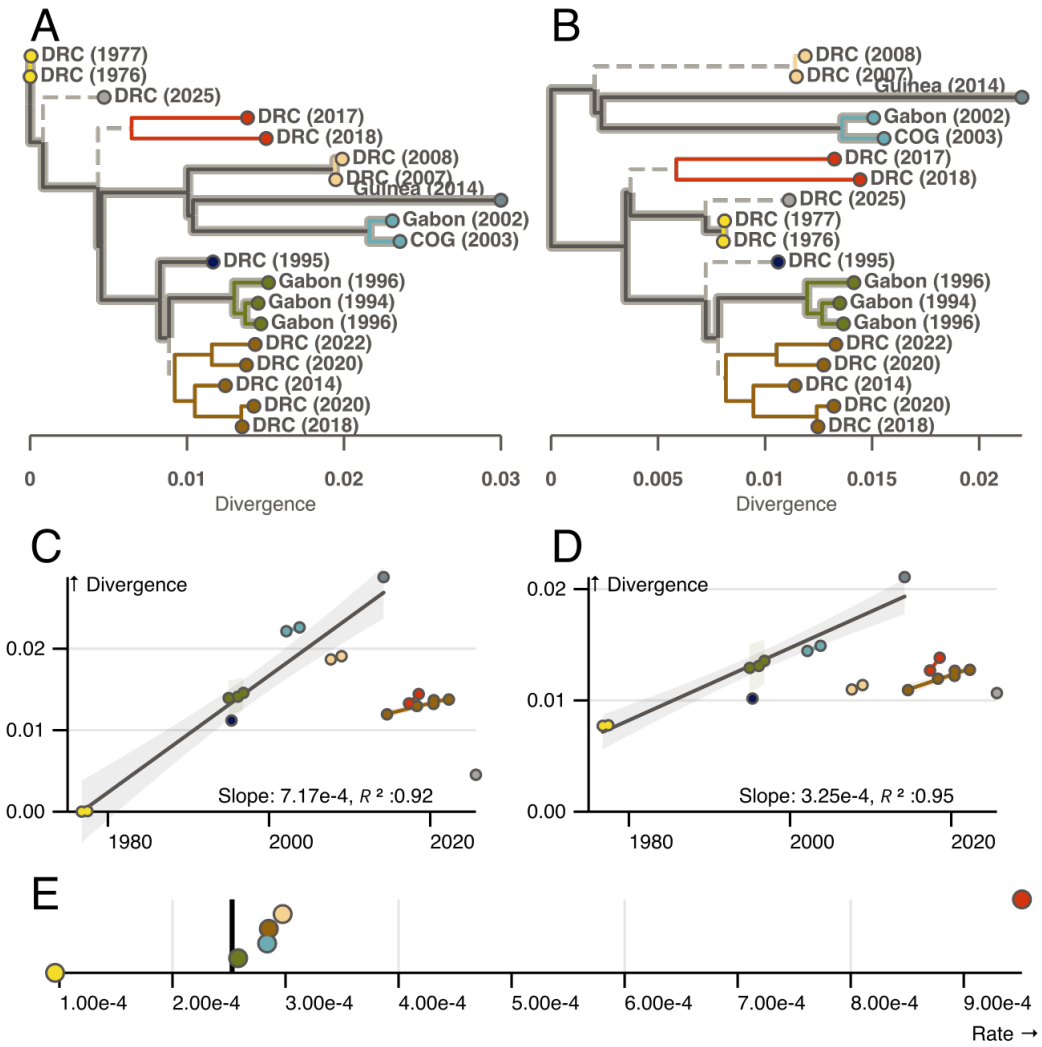

Figure 1: Root-to-tip exploration of rate heterogeneity in the EBOV phylogeny. A & C) The EBOV phylogeny rooted at the accepted 1970s outbreak position and corresponding root-to-tip divergence plot. B & D) Similar to A and C but with the proposed rooting. In all figures tips are coloured by outbreak cluster (Supplemental Figure S1A) and labeled by country and year of outbreak. Branches in A and B that are highlighted in grey connect tips used in the regressions in C and D. Smaller, coloured regression connect tips belonging to the same cluster. E) The maximum-likelihood evolutionary rate estimated for each cluster independently. The black line represents the evolutionary rate estimate when all clusters share the same rate.

We estimated the evolutionary rate of each cluster in Figure 1 by maximum-likelihood and found remarkably similar rates across many clusters (Figure 1E). We observed a slower rate of evolution among the 1970s outbreaks, although our power to estimate such a rate in the absence of many mutations is low. We also found the clade representing the 2017 and 2018 Northern DRC outbreaks may exhibit a higher rate of evolution than other clades. While this is consistent with previous results (*33*), we note that this clade is characterised by long external branches which corresponds to a high variance in the estimated rate.

We next explored possible partitions of the data that would allow for similar within- and between-cluster evolutionary rates. There is no rooting that does not require root-to-tip outliers with less divergence than their date would suggest. However, if we root the EBOV phylogeny as in Figure

Data set	Model	LL	ΔLL	df	p	Base rate	Root
Temporal set	DR	–29895.9	–	–	–	–	–
8 taxa	SRDT	–29898.0	–2.1	5	0.69	3.10×10^{-4}	Proposed
	SRDT	–29900.3	–4.4	5	0.18	1.95×10^{-4}	Midpoint
	SRDT	–29906.7	–10.9	5	< 0.001	7.56×10^{-4}	1970s
	SR	–30036.9	–141.0	6	< 0.001	–	1970s
	SR	–30794.4	–898.6	6	< 0.001	–	Midpoint
	SR	–31454.9	–1559.0	6	< 0.001	–	Proposed
Classic temporal set	DR	–31462.7	–	–	–	–	–
11 taxa + Local clocks	localSRDT	–31466.6	–3.9	6	0.37	3.05×10^{-4}	Proposed
	localSRDT	–31469.8	–7.1	6	0.03	2.02×10^{-4}	Midpoint
	localSRDT	–31480.7	–18.0	6	< 0.001	7.48×10^{-4}	1970s
	localSR	–31487.9	–25.2	7	< 0.001	–	Midpoint
	localSR	–31612.6	–149.9	7	< 0.001	–	1970s
	localSR	–38129.7	–6667.0	7	< 0.001	–	Proposed
Full data set	DR	–35710.2	–	–	–	–	–
19 taxa + Local clocks	localSRDT	–35714.7	–4.6	11	0.71	2.99×10^{-4}	Proposed
	localSRDT	–35719.1	–8.9	11	0.11	2.19×10^{-4}	Midpoint
	localSR	–35747.8	–37.7	12	< 0.001	–	Midpoint
	localSRDT	–35748.8	–38.6	11	< 0.001	6.99×10^{-4}	1970s
	localSR	–35763.3	–53.1	12	< 0.001	–	Proposed
	localSR	–36017.2	–307.1	12	< 0.001	–	1970s

Table 1: Results of a TipDate analysis (*35*) for temporal signal in the 8 outbreaks highlighted in Figure 1B and subsequent partitions with local clocks placed on non-cluster branches subtending the remaining outbreak clusters. Model - the clock model used (DR: different rates, i.e. no clock; SR: single rate, i.e. unidentifiable clock; SRDT: single rate dated tips, i.e. identifiable clock; localSRDT - an SRDT model with local clock on between cluster branches not highlighted in Figure 1B), LL - Log likelihood, ΔLL - Log likelihood difference from the best model, df - degrees of freedom in a chi-squared test, p - p-value of a chisquared test, Base rate - the inferred evolutionary rate (where applicable), Root - the root position used in the analysis (see Supplemental Figure S1B)

1B, there is a subset of branches that pass through the root, connect several outbreak clusters, and imply a similar between- and within- cluster evolutionary rate.

We found statistical support for clock-like evolution with an identifiable rate in the subtree of eight outbreaks highlighted in Figure 1B when rooted at this proposed root (*35*). The proposed root position had the highest log-likelihood of all tested rootings ("Proposed" in Table 1), but temporal signal was also observed when the tree was midpoint rooted with the 2013 West African outbreak as an outgroup. Interestingly, we found no evidence for clock-like evolution in this dataset with the accepted root position near the 1976 Yambuku outbreak. These signals were maintained in the 1976-2013 dataset ("Classic temporal set") as well as in the full dataset of 19 samples when local clocks were applied to branches subtending the remaining clades (dashed branches in Figure 1B; Table 1). Our findings suggest that while the current model of EBOV is likely based on a misguided interpretation of root-to-tip correlations, clock-like evolution with periodic rate heterogeneity is a plausible explanation for EBOV evolutionary dynamics.

A mechanistic rate model discovers the same evolutionary dynamics

We sought to determine if our proposed model for EBOV evolution is consistent with latency in the reservoir population. To this end, we developed a state-dependent evolutionary rate model in which

lineages transition between latent and replicating states (see Methods). When in the replicating state, substitutions accumulate according to an underlying evolutionary rate; however, the rate is zero whilst in the latent state. We constrain our model such that all nodes are in the replicating state. External nodes represent sampled, active infections and internal nodes represent bifurcation events - both processes that require that the virus was replicating. Because of this constraint, latency is less likely to be observed on branches that span short periods of time, such as those seen within outbreak clusters, and more likely on the longer branches that connect outbreaks.

We initially focused on the 18 outbreaks identified in Central Africa, based on the assumption that these may be connected by a contiguous reservoir population that excludes the geographically distant 2013 West African outbreak. We found the proposed root had the highest posterior support (74.4%) with the remaining support split between trees with the 2002/2003 Gabon clade (blue) or the 2007/2008 DRC clade (peach) as outgroups. (Supplemental Figure S2) This analysis resulted in a mean root age of 1931 (95% HPD [1892, 1956]) which corresponded to a mean replicating evolutionary rate of 2.34×10^{-4} s/s/y (95% HPD [1.43×10^{-4}, 3.19×10^{-4}]). We estimated a mean latent-replicating transition rate of 0.07 (95% HPD [0.03, 0.10]) transitions per year with a mean bias of 0.70 (95% HPD [0.429, 0.951]) towards entering the latent state. These parameters result in a median of 6 branches with latency (95% HPD [5, 9]). The root position, root age, and evolutionary rate are all consistent with our exploration above. Further, the five branches observed to be outliers in the root-to-tip plots are the only branches with a greater than 50% posterior probability of at least one latent period.

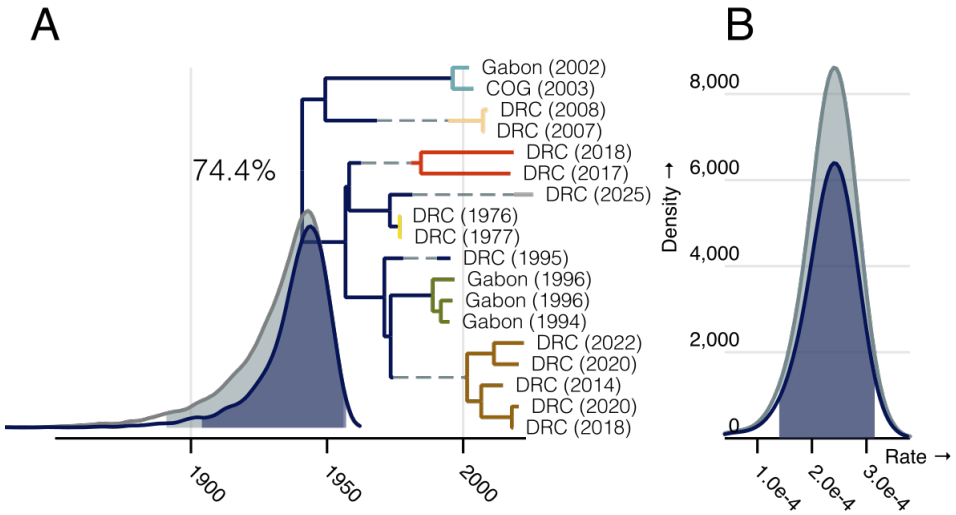

Figure 2: Posterior distributions of latency model parameter estimates when the West African outbreak is excluded. A) A maximum clade credibility (MCC) tree with the posterior probability of the root position noted near the root. Clades are coloured as previously. The root age distribution of the entire analysis is shown in grey with the contribution of the displayed root highlighted in blue. The mean duration of latency (conditioned on there being at least one period of latency) is shown as a dashed line on branches with a posterior probability of any latency greater than 50%. B) The posterior distribution of the evolutionary rate during replication, again shaded by total (grey) and root-specific conditioning (blue). In all posterior distribution plots 95% HPD intervals are shaded.

Despite being a geographic outlier, 2013 West African outbreak is not a temporal outlier (Figure 1D). The large number of substitutions that distinguish this outbreak from the other EBOV isolates are consistent with our estimated rooting and evolutionary rate (Figure 1) and suggests little to no latency has occurred on this branch. We wanted to explore whether or not a branch of this length is consistent with our estimates of latency. If we assume this outbreak diverges from the rest of EBOV near the root estimated above, its expected branch length would be 83 years long (95% HPD [58, 122]) which corresponds to a 5.61% probability (0.84%, 18.50%) of no latency.

We found our model was largely robust to inclusion of this possible outlier, with the exception of

increased uncertainty in the EBOV root placement. The mean root age of the entire dataset was 1922 (95% HPD [1884, 1954]) which corresponded to a replicating evolutionary rate of 2.37×10^{-4} substitutions per site per year (95% HPD [1.51×10^{-4}, 3.12×10^{-4}]) and a median of 7 branches with latency (95% HPD [5, 10]). The underlying evolutionary rate is consistent with that expected from the root-to-tip analysis and maximum-likelihood estimates above, but we find less support for the proposed root position (19.7% posterior support). Instead the models favours a root with the West African outbreak as an outgroup (27.4%). The majority of the remaining posterior support was again split between trees with the 2002/2003 Gabon clade (blue) or the 2007/2008 DRC clade (peach) as outgroups (Supplemental Figure S3).

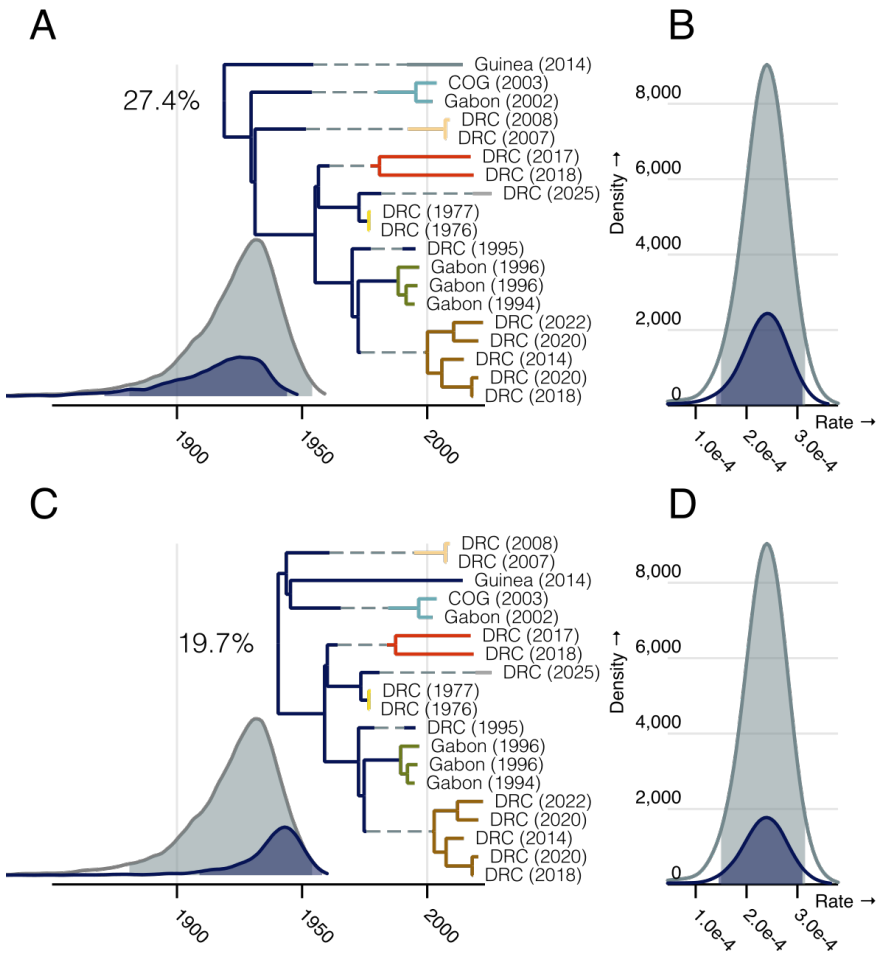

Figure 3: Posterior distributions of latency model parameters estimated for the full dataset partitioned by root position. A & C) MCC trees of the most common root placement (A) and the proposed root (C) (posterior probability noted near root). Clades are coloured as previously. The marginal posterior root age distribution is shown in grey with each root's conditional contribution is highlighted in blue. The mean duration of latency (conditioned on there being at least one period of latency) is shown as a dashed line on branches with a posterior probability of any latency greater than 50%. B & D) The posterior distribution of the evolutionary rate during replication, again shaded by total analysis (grey) and conditioning on specific roots (blue). In all posterior distribution plots the shaded area represents the 95% HPD.

Recent work by Gao *et al.* (*36*) has shown that posterior distributions can be disproportionately affected by the inclusion of a small number of taxa. Our estimate of the EBOV root position is largely affected by the inclusion of the 2013 West African outbreak. If this outbreak represents the true EBOV outgroup, the root age is likely much older than that estimated by our model, as the duration of latency on this branch (and its sibling) is only informed by the parameter estimates from other parts of the tree. However, we are able to estimate the root of the remaining 18 outbreaks with reasonable

confidence. While the rootings in Supplemental Figure S3 represent 4 distinct topologies, they all lie near our proposed root, where several long branches diverge in rapid succession (Supplemental Figure S1B). The uncertainty in the root position is defined entirely by uncertainty in the order of these events, leading us to conclude the proposed root is the most probable EBOV root position.

Consequences of latency for future outbreaks

Our results suggest a significant number of EBOV outbreaks have derived from once latent lineages, and that latency may be an evolutionary mechanism by which EBOV persists in the reservoir. We were interested in the contribution of latency to future EBOV outbreaks, and plotted a cumulative distribution of there being at least one latent period as a function of branch length using the parameters estimated by our model (Figure 4A). We excluded the West African outbreak in these parameter estimates and relied on the analysis presented in Figure 2 on the basis that the West African outbreak seems to have originated from a separate process than that acting in Central Africa. Our results suggest a branch length of 35 years has a 50% chance of exhibiting latency, and the expected proportion of time spent latent for such a branch is 0.4 or roughly 14 years. As highlighted in Figure 4B, we have little precision in our estimate of how much latency to expect provided there is latency on a branch.

Our model above, finds latency on the branch subtending the recent EBOV outbreak in Kasai, DRC. This is consistent with the limited number of mutations (96) that have accumulated over the 49 years that have passed between this outbreak its closest relative — the 1976 Yambuku outbreak. We wondered if our model estimates could be used to predict the contribution of latency to this outbreak prior to genome sequencing. Fitting the latent model to the Central African dataset without the Kasaï outbreak resulted in similar parameter estimates to those reported above: mean transition rate of 0.07 (95% HPD [0.03, 0.10]) per year with a mean bias of 0.67 (95% HPD [0.373, 0.947]) towards entering the latent state. These estimates predict the Kasaï outbreak has an expected probability of 61.8% of at least one latent period. As expected, conditional on there being latency, the model provides a wide range of possible durations. The expected time spent latent is 20 years (95% quantiles [6, 48]), which corresponds to an evolutionary rate of 1.35×10^{-4} s/s/y (95% quantiles [3.20×10^{-5}, 2.51×10^{-4}]) or roughly 134 mutations (95% quantiles [28, 218]). The observed 96 mutations falls within this range; however we expect a more detailed model that accounts for the fragmented nature of the EBOV reservoir may better predict which clades are likely replicating, and which may be expected to persist only through latency.

Discussion

An accurate model of EBOV evolution in the reservoir has important consequences for how we classify observed outbreaks and evaluate the risk of future epidemics. Following the Guinea outbreak of 2021, which was linked to the 2013 West African outbreak, it was proposed that decreased rates of evolution could be used to identify human-derived outbreaks (*37*). In the initial phylogenetic analysis of the Equateur/2020 outbreak in the DRC one genome was found to cluster near the base of the Equateur/2018 outbreak (*23*). Because of this placement, and the lack of rapid evolution, the sample was thought to represent a persistent human case from the Equateur/2018 outbreak. However, there was no epidemiological evidence for this noted at the time, and this sequence's divergence is consistent with our modified estimate of Ebola reservoir dynamics (Note the two DRC (2020) isolates in Figure 1). Given our updated model, this sample likely represents an independent spillover event from a large, diverse reservoir active in Equateur province, DRC, between 2018-2022, and not a persistent human case.

Our updated estimate of the evolutionary rate of EBOV in the reservoir has additional implications for our understanding of EBOV evolutionary dynamics. The evolutionary rate of the 2013 West African outbreak was initially observed to be slightly higher than the prevailing estimate of

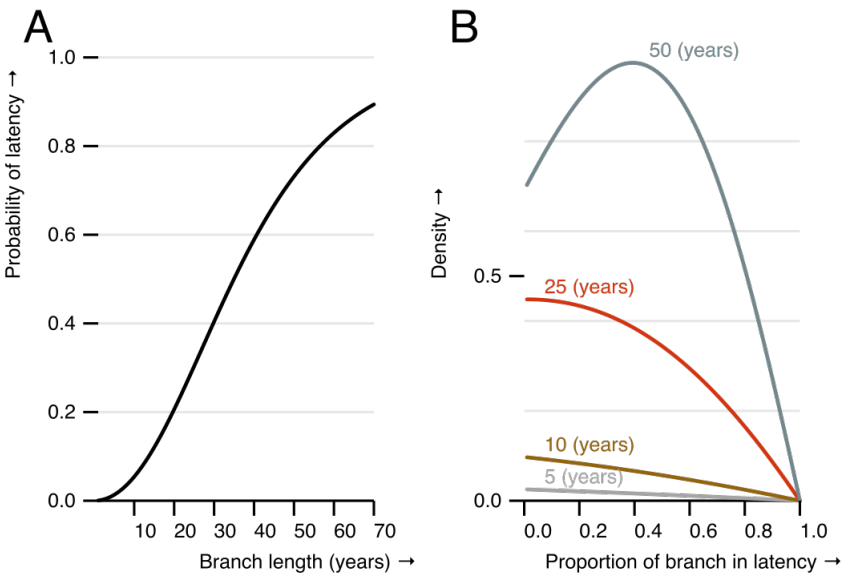

Figure 4: The expected contribution of latency to future outbreaks based on model estimates. A) The probability of some latency as a function of branch length. B) The distribution of the proportion of time spent in latency provided there is some latency on branches of different length (noted text near each line).

the reservoir rate (*38*, *39*). This is expected as decreased purifying selection is known to elevate evolutionary rate estimates taken during acute outbreaks (*40*). These estimates are consistent with a single EBOV evolutionary rate in humans and the reservoir. However, our estimated EBOV evolutionary rate is less than half of previous estimates, suggesting EBOV evolution may evolve by host-specific evolutionary processes (i.e. with elevated rates in human populations).

Given our finding that latency is wide-spread in EBOV evolution, it may be tempting to conclude that many EBOV outbreaks have resulted from persistently infected survivors of previous, unobserved outbreaks. Perhaps humans make up a significant proportion of the reservoir. We believe this to be implausible. Persistence in humans, although observed, is exceedingly rare, and only associated with the largest outbreaks and epidemics. Such an EBOV reservoir would require large, unobserved human epidemics. Additionally, our analysis identifies latency on several internal branches, precluding the possibility that these branches represent persistent human infections.

Our proposed root yields an older date for the common ancestor EBOV diversity and has important implications for understanding the geographic distribution of the virus. The older root provides more time for EBOV to have expanded into its wide geographic range, while maintaining the possibility of wave-like spread in a subclade of the tree. Our model also suggests EBOV diversity predates the first observed outbreaks in the 1970s, and does not require EBOV to have emerged from a genetic bottleneck just prior to the 1976 Yambuku outbreak. Although seemingly supported by the root-to-tip model at the time, this bottleneck and subsequent rapid expansion across the continent defied ready explanation from a disease ecology point of view.

Finally, we hypothesize that the phenomenon of EBOV persistence and latency in bat species is a mechanism by which the virus can maintain itself long term in a highly-structured host population - i.e., high-density but geographically dispersed roosts. Younger bats dispersing to other colonies for mating, whilst persistently infected, may be the route of transmission between roosts. Furthermore, persistent infections within a roost may spark a new outbreak years later after sufficient immunologically-naïve individuals have been born. This then establishes a new cohort of persistently infected individuals. Persistence combined with periods of latency between episodes of replication and transmission is consistent with the phylogenetic patterns we observe between human outbreaks.

The reservoir composition and dynamics of Ebola virus have evaded characterisation since the virus was first isolated in the late 1970s. Our work here has used phylogenetic analyses to provide

a rough sketch of the reservoir from its periodic spillover into humans. Filling in the details, the mechanisms behind latency, its implications for host dynamics, and its interaction with ecological processes like habitat fracturing, are crucial to understanding when and where EBOV will emerge next.

Methods

Data

To explore EBOV evolutionary rate variation in non-human hosts, we assembled a data set of genomes that span the known history of the virus. Most available EBOV genomes have been sampled from human cases. We have included one genome per outbreak, preferring those with precise dates of sampling. A list of sequences used is provided in Supplementary Table S2, including, where applicable, their GenBank or Pathoplexus accession numbers.

Phylogenetic inference

Sequences were aligned using MAFFT (*41*) and refined by hand. Five prime and three prime untranslated regions were ignored in all phylogenetic analyses.

In each analysis, we used a separate HKY substitution model (*42*) for each of three partitions of the data: coding position one, coding position two, and a concatenation of coding position three and the intergenic regions. This was deemed to be the best partition scheme by IQ-TREE v2's partition finder when compared to a partitioning scheme with separate partitions for the third coding position and intergenic regions. Maximum-likelihood phylogenetic trees were estimated using IQ-TREE v2.4.0(*43*). The likelihood ratio tests in the TipDate analysis (*35*) were performed using PAML (commit:ac9b97c8c35) (*44*, *45*).

For the root-to-tip correlations in Figure 1, which exclude several clades, we first rooted the tree on the specified branch and then determined the location of the root on that branch based on most likely location from the TipDate analysis above. These calculations gave very similar results to those obtained from TempEst when run on trees containing just the tips included in the regressions (highlighted branches in Figure 1 A and B) (*46*).

Clade-level, maximum-likelihood evolutionary rate estimates were calculated using the strict-clock model proposed by Didelot *et al.* (*47*). Each clade was identified in and pruned out of the full maximum likelihood tree estimated by IQ-TREE v2 so that even clades with only 2 tips were rooted. We then estimated the maximum-likelihood evolutionary rate and time-scaled the branch lengths using a gamma distribution with mean equal to the variance as described in (*47*). Optimization was achieved using fmin (`https://www.npmjs.com/package/fmin`).

All Bayesian phylogenetic analyses were performed in BEAST X (*48*) using the BEAGLE high-performance computational library (*49*). For each analysis two chains were run for 500 million states and sub-sampled every 100 000 states with the first 50 000 states removed as burn-in. We employed an constant population coalescent tree prior with a log normal prior (μ:4, σ: 1) (95% interval [8, 388]) on the population size. Preliminary analyses with an exponentially growing population did not exhibit a growth rate different from 0.

We placed an continuous-time Markov chain (CTMC) reference prior on the underlying evolutionary rate (*50*), with a slight modification. The CTMC prior is approximated as a gamma distribution with shape 0.5 and rate equal to the length of the tree, which results in an expectation of 0.5 mutations at each site over the entire tree. Because our tree contains periods during which no mutations accumulate, we updated the prior so that the rate was equal to the length of the replicating part of the tree which arrives at the same expectation.

We used an additional CTMC prior on the latent transition rate (see below for details). However, because latency is a branch specific event that 'resets' at each node and because we expect there

to be some small amount of latency over the duration of the tree, we use the root height as the rate parameter. We place a uniform prior on the relative rate bias between entering and leaving the latent state. Sampling from these prior distributions results in a mean of 0.04 branches enjoying at least one period of latency with a 96% prior probability of no branches exibiting latency. We assessed convergence to the posterior distribution and proper mixing of all relevant parameters in Tracer 1.7 (*51*).

Latent-state branch rate model

We model the evolutionary rate of EBOV according to a state-dependent rate model (*52*) where branches alternate between two possible states, replicating and latent, according to a CTMC.

Each branch transitions independently between states according to an infinitesimal rate matrix Q conditioned on starting and ending in the replicating state. This condition is necessitated by the fact that each node represents either a sampled active infection (external nodes) or an unsampled branching event (internal nodes). While in the replicating state, each branch undergoes evolution according to some underlying molecular clock rate model (e.g. strict clock, uncorrelated relaxed clock, etc.); the evolutionary rate is set to 0 when in the latent state. The overall evolutionary rate then along a branch is determined by its underlying rate (μ) and the amount of time spent in the replicating (t_r) and latent (t_l) states, where $t = t_l + t_r$ is the total length of the branch in time.

Several approaches exist for estimating the amount of time a branch spends in each state. Cappello *et al.* explicitly sample the state history of each lineage as it transitions between replicating and non-replicating states in a seedbank model (*53*). While accurate, explicitly sampling complete transition histories is computationally expensive. Similarly, Lewinsohn *et al.* have developed an alternative state-dependent evolutionary rate model that marginalizes over the state history and approximates the evolutionary rate of each branch by using the expected time spent in each state (*52*). Such an approximation seems appropriate in cases where both states replicate at different rates. However, in our setting, where branches may not experience any latency the mean time spent latent is a poor representation of the possible dynamics.

We avoid sampling migration histories and instead estimate the proportion of time each branch spends in latency, integrated over all possible histories. Our approach is akin to recent approximations to the structured coalescent which also marginalize over histories (*54–57*). However, the model used here is simplified in that each branch's migration history is independent of all others, which allows for an exact solution.

We are interested in estimating the phylogenetic posterior distribution described by

$$p(G, \Theta, \Omega, \kappa|D) \propto P(D|G, \Phi, \kappa) f(\Phi|\Omega) f_T(G|\Theta) \pi(\Omega, \Theta, \kappa), \tag{1}$$

where G is our phylogenetic tree with branch lengths in years, Θ represents the parameters governing the coalescent tree prior, Ω the parameters of the evolutionary model, and κ the substitution parameters. Moreover, $P(D|G, \Phi, \kappa)$ is the observed sequence likelihood of an alignment D given the phylogenetic tree G, the collection of evolutionary rates (one for each branch) Φ, and the parameters governing the substitution model κ. Density $f_T(G|\Theta)$ is the tree prior with hyperparameters Θ, $f(\Phi|\Omega)$ the probability of each branch's evolutionary rate according to the latent model described below, and $\pi(\Omega, \Theta, \kappa)$ represents the prior over the remaining parameters.

In our model, Ω parametrizes the rate matrix underlying our latent-state CTMC and is comprised of two parameters, r and b, in addition to any parameters needed by the base evolutionary rate model – μ in the case of a strict clock. The rate matrix is defined as

$$Q = \begin{bmatrix} -rb & rb \\ r(1-b) & -r(1-b) \end{bmatrix} \tag{2}$$

with the first state representing replication and the second latency.

Given that all nodes exist in the replicate state, the process on each branch is then independent of the others and so we write

$$f(\Phi|\Omega) = \prod_{i=1}^{2(n-1)} f(\phi_i|\Omega), \tag{3}$$

where n is the number of tips in our tree, $2(n-1)$ is the number of branches and each branch has its own rate ϕ. We define each $\phi = \mu p_l$ as the product of a base evolutionary rate μ and the proportion of the branch length spent in latency p_l. Here, we use a strict clock model to assign the same base evolutionary rate to each branch, but more complicated models could be used. Thus,

$$f(\phi|\Omega, t) = f(p_l|\Omega, t), \tag{4}$$

where t is the length of the branch in time. We derive $f(p_l|\Omega, t)$ below.

We note that $p_l = t_l/t$ and so

$$f(p_l|\Omega, t) = \begin{cases} f(t_l|\Omega, t) \cdot t & 0 < p_l \leq 1, \\ f(t_l|\Omega, t) & p_l = 0. \end{cases} \tag{5}$$

Let X(s) describe the state of a branch (i.e. rep, latent) at time s ($0 \leq s \leq t$), and t_l describe the occupation time spent in the latent state. Sericola and colleagues (*58*, *59*) provide numerically stable algorithms for the cumulative distribution function

$$F(t_l \leq s, X(t) = rep|X(0) = rep, t, \Omega), \ s \in [0, t], \tag{6}$$

and probability distribution function

$$f(t_l = s, X(t) = rep|X(0) = rep, t, \Omega), \ s \in (0, t], \tag{7}$$

which are conditioned on the starting state being in the replicating state (X(0) = rep).

Here, we are interested in conditioning on both the starting (X(0)) and ending (X(t)) state and have

$$F(t_l \leq s|X(t) = rep, X(0) = rep, t, \Omega) = \frac{F(t_l \leq s, X(t) = rep|X(0) = rep, t, \Omega)}{f(X(t) = rep|X(0) = rep)} \tag{8}$$

and

$$f(t_l = s|X(t) = rep, X(0) = rep, t, \Omega) = \frac{f(t_l \leq s, X(t) = rep|X(0) = rep, t, \Omega)}{f(X(t) = rep|X(0) = rep)}, \tag{9}$$

which we will write with a t subscript below ($F_t(t_l|\Omega, t)$ and $f_t(t_l|\Omega, t)$) to denote they are functions of the total branch length t. The probability of never entering the latent state $t_l = 0$ is given by $F(0|\Omega, t)$. The probability distribution function of the proportion of time spent latent $f(p_l|\Omega, t)$ is then defined as

$$f(p_l|\Omega, t) = \begin{cases} t \cdot f_t(p_l \cdot t|\Omega, t) & 0 < p_l \leq 1, \\ F_t(0|\Omega, t) & p_l = 0. \end{cases} \tag{10}$$

The latent-state model has been implemented in BEAST X (commit:edf73f) (*48*). In our analyses we place a uniform $(0, 1)$ prior on b and modified CTMC prior with rate equal to the height of the root on on r.

Competing interests

None declared.

Acknowledgments

JTM, MAS and AR are grateful for support from the Wellcome Trust (Collaborators Award 206298/Z/17/Z, ARTIC network). GB acknowledges support from the Research Foundation - Flanders ("Fonds voor Wetenschappelijk Onderzoek - Vlaanderen," G0E1420N, G098321N), from the European Union Horizon 2023 RIA project LEAPS (grant agreement no. 101094685), and from the DURABLE EU4Health project 02/2023-01/2027, which is co-funded by the European Union (call EU4H-2021-PJ4) under Grant Agreement No. 101102733. LMC was partly funded by FAPERJ - Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Processo SEI 260003/005679/2023 and SEI 260003/013252/2024. GD acknowledges the support of European Molecular Biology Organization (EMBO) installation grant IG-5305-2023. PM and AR acknowledge support from the Wellcome Trust (Discretionary Award 313694/Z/24/Z, ARTIC 2). IFO thanks the Wellcome Trust Hosts, Pathogens & Global Health program (Wellcome Trust, grant 218471/Z/19/Z) in partnership with Tackling Infectious Disease to Benefit Africa. MAS and AR acknowledge support from National Institutes of Health (NIH) grants R01 AI153044 and R01 AI162611 and the European Research Council (grant agreement no. 725422 – ReservoirDOCS).

References

1. I. Commission, en, *Bull. World Health Organ.* **56**, 271–293 (1978).
2. en, *Bull. World Health Organ.* **56**, 247–270 (1978).
3. J. S. Towner *et al.*, en, *PLoS Pathog.* **5**, e1000536 (July 2009).
4. J. T. Pawęska *et al.*, en, *Emerg. Infect. Dis.* **26**, 3051–3055 (Dec. 2020).
5. E. M. Leroy *et al.*, en, *Nature* **438**, 575–576 (Dec. 2005).
6. K. Kupferschmidt, *This bat species may be the source of the Ebola epidemic that killed more than 11,000 people in West Africa*, en, https://www.science.org/content/article/bat-species-may-be-source-ebola-epidemic-killed-more-11000-people-west-africa, Accessed: 2025-9-11, Mar. 2021.
7. R. Kessler, *EcoHealth Alliance Scientists Discover the Deadly Zaire Ebola Virus in West African Bat*, en, https://www.ecohealthalliance.org/2019/01/ecohealth-alliance-scientists-discover-the-deadly-zaire-ebola-virus-in-west-african-bat, Accessed: 2025-9-11, Jan. 2019.
8. R. Kessler, *Chasing Ebola to Its Source*, en, https://www.ecohealthalliance.org/2019/02/chasing-ebola-to-its-source, Accessed: 2025-9-11, Feb. 2019.
9. E. M. Leroy *et al.*, en, *Science* **303**, 387–390 (Jan. 2004).
10. E. M. Leroy *et al.*, en, *J. Infect. Dis.* **190**, 1895–1899 (Dec. 2004).
11. T. J. Wittmann *et al.*, en, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17123–17127 (Oct. 2007).
12. A. Lacroix *et al.*, en, *Pathogens* **10**, 557 (May 2021).
13. R. Biek, P. D. Walsh, E. M. Leroy, L. A. Real, en, *PLoS Pathog.* **2**, e90 (Oct. 2006).
14. P. D. Walsh, R. Biek, L. A. Real, en, *PLoS Biol.* **3**, e371 (Nov. 2005).
15. D. L. Heymann *et al.*, en, *J. Infect. Dis.* **142**, 372–376 (Sept. 1980).
16. S. A. Carroll *et al.*, *J. Virol.* **87**, 2608–2616 (Feb. 2013).
17. G. Grard *et al.*, en, *J. Infect. Dis.* **204 Suppl 3**, S776–84 (Nov. 2011).
18. S. Baize *et al.*, en, *N. Engl. J. Med.* **371**, 1418–1425 (Oct. 2014).
19. W. C. Wheeler, *Cladistics* **6**, 363–367, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1096-0031.1990.tb00550.x, (https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1096-0031.1990.tb00550.x) (1990).
20. G. Dudas, A. Rambaut, en, *PLoS Curr.* **6** (May 2014).
21. T. T.-Y. Lam, H. Zhu, Y. L. Chong, E. C. Holmes, Y. Guan, en, *J. Virol.* **89**, 10130–10132 (Oct. 2015).
22. J. Nsio *et al.*, en, *J. Infect. Dis.* **221**, 701–706 (Feb. 2020).
23. E. Kinganda-Lusamaki *et al.*, en, *Lancet Microbe* (Jan. 2024).
24. P. Mbala-Kingebeni *et al.*, en, *Lancet Infect. Dis.* **19**, 641–647 (June 2019).
25. *April 2022 Ebola virus disease case in Equateur Province, DRC, represents a new spillover event*, en, https://virological.org/t/april-2022-ebola-virus-disease-case-in-equateur-province-drc-represents-a-new-spillover-event/795, Accessed: 2025-2-21.

26. *The 16th Ebola Virus Disease Outbreak in Bulape Health Zone, Kasai, Democratic Republic of the Congo: A new spillover event from an unknown reservoir host*, en, https://virological.org/t/the-16th-ebola-virus-disease-outbreak-in-bulape-health-zone-kasai-democratic-republic-of-the-congo-a-new-spillover-event-from-an-unknown-reservoir-host/1003, Accessed: 2025-10-15, Sept. 2025.
27. S. L. M. Whitmer *et al.*, en, *Cell Rep.* **22**, 1159–1168 (Jan. 2018).
28. B. Diallo *et al.*, en, *Clin. Infect. Dis.* **63**, 1353–1356 (Nov. 2016).
29. A. E. Thorson *et al.*, en, *PLoS Med.* **18**, e1003273 (Feb. 2021).
30. A. K. Keita *et al.*, en, *Nature* **597**, 539–543 (Sept. 2021).
31. M. Jacobs *et al.*, en, *Lancet* **388**, 498–503 (July 2016).
32. E. C. Holmes, G. Dudas, A. Rambaut, K. G. Andersen, en, *Nature* **538**, 193–200 (Oct. 2016).
33. P. Mbala-Kingebeni *et al.*, en, *Lancet Infect. Dis.* **19**, 648–657 (June 2019).
34. /. A. Rambaut, J. T. McCrone, G. Baele, *Ebola Virus Local Clock Analysis*, https://beast.community/ebov_local_clocks.html, Accessed: 2025-2-21, May 2019.
35. A. Rambaut, en, *Bioinformatics* **16**, 395–399 (Apr. 2000).
36. J. Gao *et al.*, *bioRxiv*, eprint: https://www.biorxiv.org/content/early/2025/06/12/2025.06.10.657742.full.pdf, (https://www.biorxiv.org/content/early/2025/06/12/2025.06.10.657742) (2025).
37. R. F. Garry, en, *Nature* **597**, 478–480 (Sept. 2021).
38. S. K. Gire *et al.*, en, *Science* **345**, 1369–1372 (Sept. 2014).
39. D. J. Park *et al.*, en, *Cell* **161**, 1516–1526 (June 2015).
40. M. Ghafari *et al.*, en, *Mol. Biol. Evol.* **39**, msac009 (Feb. 2022).
41. K. Katoh, D. M. Standley, en, *Mol. Biol. Evol.* **30**, 772–780 (Apr. 2013).
42. M. Hasegawa, H. Kishino, T. Yano, en, *J. Mol. Evol.* **22**, 160–174 (1985).
43. B. Q. Minh *et al.*, en, *Mol. Biol. Evol.* **37**, 1530–1534 (May 2020).
44. Z. Yang, en, *Mol. Biol. Evol.* **24**, 1586–1591 (Aug. 2007).
45. Z. Yang, en, *Comput. Appl. Biosci.* **13**, 555–556 (Oct. 1997).
46. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, en, *Virus Evol* **2**, vew007 (Jan. 2016).
47. X. Didelot, N. J. Croucher, S. D. Bentley, S. R. Harris, D. J. Wilson, en, *Nucleic Acids Res.* **46**, e134 (Dec. 2018).
48. G. Baele *et al.*, en, *Nat. Methods* **22**, 1653–1656 (Aug. 2025).
49. D. L. Ayres *et al.*, *Systematic biology* **68**, 1052–1061 (2019).
50. M. A. R. Ferreira, M. A. Suchard, *Canadian Journal of Statistics* **36**, 355–368, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjs.5550360302, (https://onlinelibrary.wiley.com/doi/abs/10.1002/cjs.5550360302) (2008).
51. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, *Systematic Biology* **67**, 901–904, ISSN: 1063-5157, eprint: https://academic.oup.com/sysbio/article-pdf/67/5/901/25517397/syy032.pdf, (https://doi.org/10.1093/sysbio/syy032) (Apr. 2018).
52. M. A. Lewinsohn, T. Bedford, N. F. Müller, A. F. Feder, en, *Nat Ecol Evol* **7**, 581–596 (Apr. 2023).
53. L. Cappello *et al.*, en, *bioRxivorg*, 2025.01.19.633741 (Jan. 2025).
54. N. F. Müller, D. Rasmussen, T. Stadler, en, *Bioinformatics* **34**, 3843–3848 (Nov. 2018).

55. N. F. Müller, D. A. Rasmussen, T. Stadler, en, *Mol. Biol. Evol.* **34**, 2970–2981 (Nov. 2017).

56. N. De Maio, C.-H. Wu, K. M. O'Reilly, D. Wilson, en, *PLoS Genet.* **11**, e1005421 (Aug. 2015).

57. Y. Shao *et al.*, *bioRxiv*, eprint: `https://www.biorxiv.org/content/early/2025/09/24/2025.09.22.677844.full.pdf`, (`https://www.biorxiv.org/content/early/2025/09/24/2025.09.22.677844`) (2025).

58. B. Sericola, en, *Commun. Stat.* **16**, 479–510 (Jan. 2000).

59. M. Bladt, B. Meini, M. F. Neuts, B. Sericola, presented at the Matrix-Analytic Methods.

Supplemental figures and text

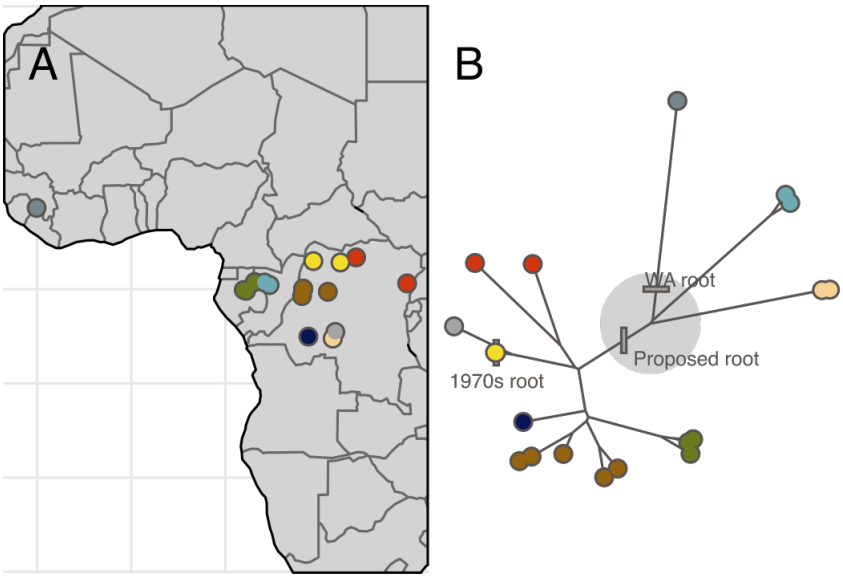

Figure S1: A map (A) and unrooted tree (B) representing the outbreaks included in this study. Common and proposed root placements are noted on the tree. The large grey sphere represents plausible root positions proposed by the latency rate model.

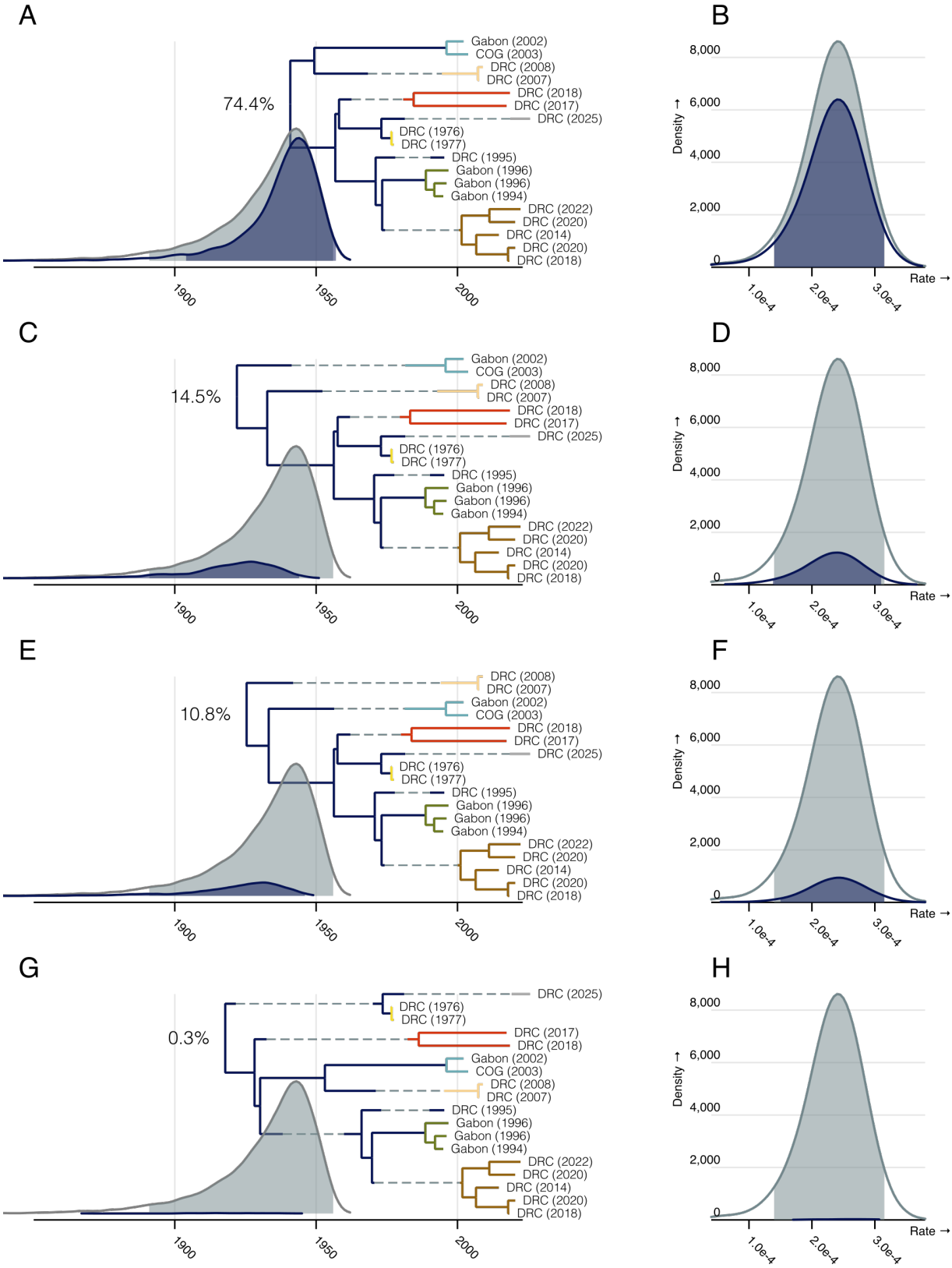

Figure S2: Posterior distributions of latency model parameters partitioned by root position for the four most common rootings in the Central Afrian data set (i.e. no West African outbreak). A, C, E, & G) MCC trees of the most common root placements (posterior probabilities are noted near each root). Clades are coloured as previously. The marginal posterior root age distribution is shown in grey with each root's conditional contribution highlighted in blue. The mean duration of latency (conditioned on there being at least one period of latency) is shown as a dashed line on branches with a posterior probability of any latency greater than 50%. B, D, F, & H) The posterior distribution of the evolutionary rate during replication, again shaded by total analysis (grey) and conditioning on specific roots (blue). In all posterior distribution plots the shaded area represents the 95% HPD.

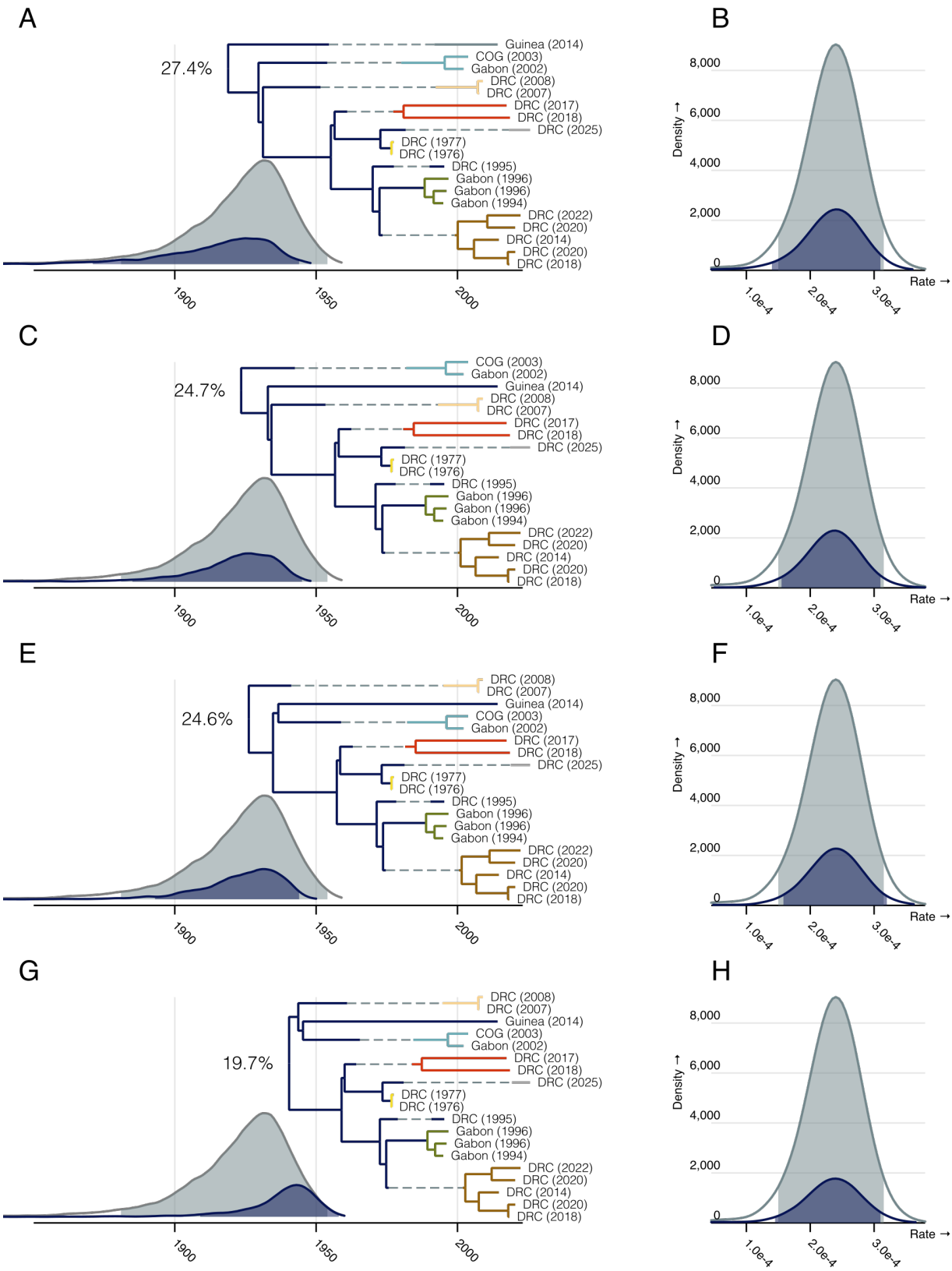

Figure S3: Posterior distributions of latency model parameters partitioned by root position for the four most common rootings in the full data set. A, C, E, & G) MCC trees of the most common root placements(posterior probabilities are noted near each root). Clades are colored as previously. The marginal posterior root age distribution is shown in grey with each root's conditional contribution highlighted in blue. The mean duration of latency (conditioned on there being at least one period of latency) is shown as a dashed line on branches with a posterior probability of any latency greater than 50%. B, D, F, & H) The posterior distribution of the evolutionary rate during replication, again shaded by total analysis (grey) and conditioning on specific roots (blue). In all posterior distribution plots the shaded area represents the 95% HPD.

outbreak	country	adm1	dates
Yambuku/1976	DRC	Mongala	Sep-Nov 1976
Sud-Ubangi/1977	DRC	Sud-Ubangi	Jun-77
Gabon/1994	GAB	Ogooue-Ivindo	Dec 1994-Feb 1995
Kwilu/1995	DRC	Kwilu	Jan-Jul 1995
Gabon/1996a	GAB	Ogooue-Ivindo	Jan–Apr 1996
Gabon/1996b	GAB	Ogooue-Ivindo	Jul-96
Gabon/2001	GAB	Ogooue-Ivindo	Oct 2001-Jul 2002
RoC/2003	COG	Cuvette Ouest	Oct-Dec 2003
Kasaï/2007	DRC	Kasai Occidental	May-Nov 2007
Kasaï/2008	DRC	Kasai Occidental	Dec 2008–Feb 2009
Makona/2013	GIN	Gueckedou	Dec 2013-Mar 2016
Tshuapa/2014	DRC	Tshuapa	Aug–Nov 2014
Bas-Uele/2017	DRC	Bas Uele	May-July 2017
Equateur/2018	DRC	Equateur	May-July 2018
Kivu/2018	DRC	Nord Kivu - Ituri - South Kiev	Aug 2018-June 2020
Equateur/2020	DRC	Equateur	June - Nov 2020
Mbandaka/2022	DRC	Equateur	April- July 2022
Kasaï/2025	DRC	Kasai Occidental	August - ongoing

Table S1: Recorded EBOV outbreaks used in this study. amd1 - Administrative Region 1 of the outbreak.

accession	country	outbreak	date	authors
KR063671	DRC	Yambuku/1976	1976-10-01	Das,S.R., Shabman,R., Halpin,R.A., Lin,X., Ransier,A., Fedorova,N., Tsitrin,T., Puri,V., Stockwell,T., Amedeo,P., Bishop,B., Gupta,N., Katzel,D., Schobel,S., Shrivastava,S., Alfson,K., Wentworth,D.E. and Griffiths,A.
KC242791	DRC	Sud-Ubangi/1977	1977-06	Carroll,S.A., Towner,J.S., Sealy,T.K., McMullan,L.K., Khristova,M.L., Burt,F.J., Swanepoel,R., Rollin,P.E. and Nichol,S.T.
KC242792	GAB	Gabon/1994	1994-12-27	Carroll,S.A., Towner,J.S., Sealy,T.K., McMullan,L.K., Khristova,M.L., Burt,F.J., Swanepoel,R., Rollin,P.E. and Nichol,S.T.
KU182905	DRC	Kwilu/1995	1995-05-04	Das,S.R., Shabman,R., Halpin,R.A., Akopov,A., Fedorova,N., Puri,V., Stockwell,T., Amedeo,P., Bishop,B., Katzel,D., Schobel,S., Shrivastava,S., Brasel,T., Yun,N., Paessler,S., Dowling,W. and Barrett,A.
KC242793	GAB	Gabon/1996a	1996-02	Carroll,S.A., Towner,J.S., Sealy,T.K., McMullan,L.K., Khristova,M.L., Burt,F.J., Swanepoel,R., Rollin,P.E. and Nichol,S.T.
KC242798	GAB	Gabon/1996b	1996-10-27	Carroll,S.A., Towner,J.S., Sealy,T.K., McMullan,L.K., Khristova,M.L., Burt,F.J., Swanepoel,R., Rollin,P.E. and Nichol,S.T.
KC242800	GAB	Gabon/2001	2002-02-23	Carroll,S.A., Towner,J.S., Sealy,T.K., McMullan,L.K., Khristova,M.L., Burt,F.J., Swanepoel,R., Rollin,P.E. and Nichol,S.T.
KF113529	COG	RoC/2003	2003-10	Chiu,C.Y., Naccache,S.N., Fair,J.N., Schneider,B.S. and Grard,G.
HQ613403	DRC	Kasaï/2007	2007-08-31	Grard,G., Biek,R., Muyembe-Tamfum,J.-J., Formenty,P., Paweska,J. and Leroy,E.
HQ613402	DRC	Kasaï/2008	2008-12-31	Grard,G., Biek,R., Muyembe-Tamfum,J.-J., Formenty,P., Paweska,J. and Leroy,E.
KJ660347	GIN	Makona/2013	2014-03-20	Baize,S., Pannetier,D., Oestereich,L., Rieger,T., Koivogui,L., Magassouba,N., Soropogui,B., Sow,M.S., Keita,S., De Clerck,H., Tiffany,A., Dominguez,G., Loua,M., Traore,A., Kolie,M., Malano,E.R., Heleze,E., Bocquin,A., Mely,S., Raoul,H., Caro,V., Cadar,D., Gabriel,M., Pahlmann,M., Tappe,D., Schmidt-Chanasit,J., Impouma,B., Diallo,A.K., Formenty,P., Van Herp,M. and Gunther,S.
KP271018	DRC	Tshuapa/2014	2014-08-20	Naccache,S.N., Mbala,P., Ngay,I., Makuwa,M., Mulembakani,P., Muyembe,J.J., Schneider,B.S. and Chiu,C.Y.
MH613311	DRC	Bas-Uele/2017	2017-05-07	Nsio,J., Kapetshi,J., Makiala,S., Formenty,P., Raymond,F., Tshapenda,G., Boucher,N., Corbeil,J., Okitandjate,A., Mbuyi,G., Kiyele,M., Mondonge,V., Kikoo,M.J., Van Herp,M., Rollin,P., Barboza,P., Muyembe Muzinga,B., Kalenga,O.I., Ahuka,S., Fausther-Bovendo,H., Kebela Ilunga,B., Kobinger,G.P. and Muyembe,J.-J.T.
MH733477	DRC	Equateur/2018	2018-05-10	Mbala,P., Pratt,C., Wiley,M.R., Makiala-Mandanda,S., Aziza,A., Di Paola,N., Diagne,M.M., Chitty,J.A., Diop,M., Ayouba,A., Vidal,N., Faye,O., Karhemere,S., Aruna,A., Nsio,J., Mulangu,F., Mukadi,D., Mukadi,P., Kombe,J., Mulumba,A., Duraffour,S., Likofata,J., Pukuta,E., Minogue,T., Sozhamannan,S., Gross,S., Schroth,G., Delaporte,E., Sanchez-Lockhart,M., Peeters,M., Muyembe,J.-J., Alpha Sall,A., Palacios,G. and Ahuka-Mundeke,S.

accession	country	outbreak	date	authors
MK007330	DRC	Kivu/2018	2018-07-28	Mbala-Kingebeni,P., Aziza,A., Di Paola,N., Wiley,M.R., Makiala-Mandanda,S., Caviness,K., Pratt,C.B., Prieto,K., Chitty,J.A., Larson,P., Ayouba,A., Vidal,N., Karhemere,S., Diop,M., Diagne,M.M., Faye,M., Faye,O., Aruna,A., Nsio,J., Mulanga,F., Mukadi,D., Mukadi,P., Kombe,J., Mulumba,A., Duraffour,S., Likofata,J., Pukuta,E., Gonzalez,J., Bartlett,M.L., Sozhamannan,S., Gross,S., Schroth,G., Kuhn,J., Delaporte,E., Sanchez-Lockhart,M., Alpha Sall,A., Muyembe,J.-J., Peeters,M., Palacios,G. and Ahuka-Mundeke,S.
OR084849	DRC	Equateur/2020	2020-05-31	Kinganda-Lusamaki,E., Whitmer,S., Lokilo-Lofiko,E., Amuri-Aziza,A., Muyembe-Mawete,F., Makangara-Cigolo,J.C., Makaya,G., Mbuyi,F., Whitesell,A., Kallay,R., Choi,M., Pratt,C., Mukadi-Bamuleka,D., Kavunga-Membo,H., Matondo-Kuamfumu,M., Mambu-Mbika,F., Ekila-Ifinji,R., Shoemaker,T., Stewart,M., Eng,J., Rajan,A., Soke,G.N., Fonjungo,P.N., Otshudiema,J.O., Folefack,G.L.T., Pukuta-Simbu,E., Talundzic,E., Shedroff,E., Bokete,J.L., Legand,A., Formenty,P., Mores,C.N., Porzucek,A.J., Tritsch,S.R., Kombe,J., Tshapenda,G., Mulangu,F., Ayouba,A., Delaporte,E., Peeters,M., Wiley,M.R., Montgomery,J.M., Klena,J.D., Muyembe-Tamfum,J.-J., Ahuka-Mundeke,S. and Mbala-Kingebeni,P.
OR084846	DRC	Equateur/2020	2020-06-12	Kinganda-Lusamaki,E., Whitmer,S., Lokilo-Lofiko,E., Amuri-Aziza,A., Muyembe-Mawete,F., Makangara-Cigolo,J.C., Makaya,G., Mbuyi,F., Whitesell,A., Kallay,R., Choi,M., Pratt,C., Mukadi-Bamuleka,D., Kavunga-Membo,H., Matondo-Kuamfumu,M., Mambu-Mbika,F., Ekila-Ifinji,R., Shoemaker,T., Stewart,M., Eng,J., Rajan,A., Soke,G.N., Fonjungo,P.N., Otshudiema,J.O., Folefack,G.L.T., Pukuta-Simbu,E., Talundzic,E., Shedroff,E., Bokete,J.L., Legand,A., Formenty,P., Mores,C.N., Porzucek,A.J., Tritsch,S.R., Kombe,J., Tshapenda,G., Mulangu,F., Ayouba,A., Delaporte,E., Peeters,M., Wiley,M.R., Montgomery,J.M., Klena,J.D., Muyembe-Tamfum,J.-J., Ahuka-Mundeke,S. and Mbala-Kingebeni,P.
PP_003V5NQ	DRC	Equateur/2022	2022-04-25	Placide Mbala-Kingebeni, Jean-Jacques Muyembe, Steve Ahuka-Mundeke, Hugo Kavunga, Daniel Mukadi, Elisabeth Pukuta, Eddy Kinganda Lusamaki, Amuri Aziza, Jean Claude Makangara, Emmanuel Lokilo, Franck Edidi, Junior Bula Bula, Raphael Lumembe, Gabriel Kabamba, Fabrice Mambu, Joel Montgomery, Peter Fonjungo, Norbert Soke, Jacques Likofata, Vital Mondonge, Gervais Folefack, Prosper Djiguimde, John Otshudiema, Deby Mukendi, Dieudonné Mwamba, John Kombe, Sofonias Tessema, Didi Bofaka, Andrew Rambaut, Mike Wiley, Catherine Pratt

accession	country	outbreak	date	authors
PP_003RXHG	DRC	Kasaï/2025	2025-09-01	Adrienne Amuri-Aziza, Gradi Luakanda - Ndelemo, Jean-Claude Makangara-Cigolo, Prince Akil-Bandali, Princesse Paku-Tshambu, Sam Wilkinson, Louis Tshulo, Josh Quick, Andre Citenga, Chloe Muswamba-Kayembe, Olga Ntumba-Tshitenge, Emmanuel Lokilo-Lofiko, Fiston Cikaya-Kankolongo, Ola Rilia, Servet Kimbonza, Elisabeth Pukuta, Daniel Mukadi, Christian Ngandu, Mathias Mossoko, Gabriel Kabamba -Lungenyi, Raphael Lumembe-Numbi, Elzedeck Mabika-Bope, Patrick Mukadi, Catherine Pratt, Dieudonne Mwamba, Nick Loman, Andrew Rambaut, Eddy Kinganda-Lusamaki, Tony Wawina-Bokalanga, Dieudonne Mumba - Ngoyi, Jean-Jacques Muyembe-Tamfum, Steve Ahuka-Mundeke Placide Mbala-Kingebeni

Table S2: A list of the genomes used in this study with outbreak and author list.