

Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks

<https://doi.org/10.1038/s41586-021-03901-9>

Received: 6 April 2021

Accepted: 11 August 2021

Published online: 15 September 2021

 Check for updates

Alpha Kabinet Keita^{1,2,26}✉, Fara R. Koundouno^{3,4,26}, Martin Faye^{5,26}, Ariane Düx^{6,26}, Julia Hinzmann^{4,7,8,26}, Haby Diallo¹, Ahidjo Ayouba², Frederic Le Marcis^{1,2,9}, Barré Soropogui³, Kékoura Ifono^{3,4}, Moussa M. Diagne⁵, Mamadou S. Sow^{1,10}, Joseph A. Bore^{3,11}, Sebastien Calvignac-Spencer⁶, Nicole Vidal², Jacob Camara³, Mamadou B. Keita¹², Annick Renevey^{4,7}, Amadou Diallo⁵, Abdoul K. Soumah¹, Saa L. Millimono^{3,4}, Almudena Mari-Saez⁶, Mamadou Diop⁵, Ahmadou Doré³, Fodé Y. Soumah¹⁰, Kaka Kourouma¹², Nathalie J. Vielle^{4,13}, Cheikh Loucoubar⁵, Ibrahima Camara¹, Karifa Kourouma^{3,4}, Giuditta Annibaldis^{4,13}, Assaïtou Bah³, Anke Thielebein^{4,7}, Meike Pahlmann^{4,7}, Steven T. Pullan^{8,11}, Miles W. Carroll^{8,11}, Joshua Quick¹⁴, Pierre Formenty¹⁵, Anais Legand¹⁵, Karla Pietro¹⁶, Michael R. Wiley^{16,17}, Noel Tordo¹⁸, Christophe Peyrefitte⁵, John T. McCrone¹⁹, Andrew Rambaut¹⁹, Youssouf Sidibé²⁰, Mamadou D. Barry²⁰, Madeleine Kourouma²⁰, Cé D. Saouromou²⁰, Mamadou Condé²⁰, Moussa Baldé¹⁰, Moriba Povogui¹, Sakoba Keita²¹, Mandiou Diakite^{22,23}, Mamadou S. Bah²², Amadou Sidibe⁹, Dembo Diakite¹⁰, Fodé B. Sako¹⁰, Fodé A. Traore¹⁰, Georges A. Ki-Zerbo¹³, Philippe Lemey²⁴, Stephan Günther^{4,7,13}, Liana E. Kafetzopoulou^{4,7,24}, Amadou A. Sall⁵, Eric Delaporte^{2,25}, Sophie Duraffour^{4,7,13,27}, Ousmane Faye^{5,27}, Fabian H. Leendertz^{6,27}, Martine Peeters^{2,27}, Abdoulaye Toure^{1,12,27} & N'. Faly Magassouba^{3,27}

Seven years after the declaration of the first epidemic of Ebola virus disease in Guinea, the country faced a new outbreak—between 14 February and 19 June 2021—near the epicentre of the previous epidemic^{1,2}. Here we use next-generation sequencing to generate complete or near-complete genomes of *Zaire ebolavirus* from samples obtained from 12 different patients. These genomes form a well-supported phylogenetic cluster with genomes from the previous outbreak, which indicates that the new outbreak was not the result of a new spillover event from an animal reservoir. The 2021 lineage shows considerably lower divergence than would be expected during sustained human-to-human transmission, which suggests a persistent infection with reduced replication or a period of latency. The resurgence of *Zaire ebolavirus* from humans five years after the end of the previous outbreak of Ebola virus disease reinforces the need for long-term medical and social care for patients who survive the disease, to reduce the risk of re-emergence and to prevent further stigmatization.

At least 30 outbreaks of Ebola virus disease (EVD) have been identified since the late 1970s, the most severe of which affected Guinea, Sierra Leone and Liberia from December 2013 to June 2016^{1,2}. Guinea experienced a new outbreak of EVD in 2021, which started in Gouéké—a town about 200 km away from the epicentre of the 2013–2016 outbreak. The probable index case was a 51-year-old nurse, an assistant of the hospital midwife in Gouéké. On 21 January 2021, she was admitted to hospital in Gouéké suffering from headache, asthenia, nausea, anorexia, vertigo

and abdominal pain. She was diagnosed with malaria and salmonellosis and was released two days later. Feeling ill again once at home, she attended a private clinic in Nzérékoré (40 km away) and visited a traditional healer, but died three days later. In the week after her death, her husband—as well as other family members who attended her funeral—fell ill, and four of them died. They were reported as the first suspect cases by the national epidemic alert system on 11 February. On 12 February, blood was taken from two suspect cases admitted to

¹Centre de Recherche et de Formation en Infectiologie de Guinée (CERFIG), Université de Conakry, Conakry, Guinea. ²TransVIHMI, Montpellier University/IRD/INSERM, Montpellier, France.

³Laboratoire du Projet des Fièvres Hémorragiques de Guinée (PFHG), Conakry, Guinea. ⁴Bernhard Nocht Institute for Tropical Medicine (BNITM), Hamburg, Germany. ⁵Institut Pasteur de Dakar (IPD), Dakar, Senegal. ⁶Robert Koch Institute (RKI), Berlin, Germany. ⁷German Center for Infection Research (DZIF), Partner Site Hamburg–Lübeck–Borstel–Riems, Hamburg, Germany. ⁸UK Health Security Agency, National Infection Service, Porton Down, Salisbury, UK. ⁹Ecole Nationale Supérieure de Lyon, Lyon, France. ¹⁰Hôpital National Donka, Service des Maladies Infectieuses et Tropicales, Conakry, Guinea. ¹¹Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ¹²Institut National de Santé Publique de Guinée (INSP), Conakry, Guinea. ¹³World Health Organization (WHO), Conakry, Guinea. ¹⁴Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. ¹⁵Health Emergencies Program, World Health Organization (WHO), Geneva, Switzerland. ¹⁶PræsentBio, Lincoln, NE, USA. ¹⁷University of Nebraska Medical Center, Omaha, NE, USA. ¹⁸Institut Pasteur de Guinée, Conakry, Guinea. ¹⁹Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. ²⁰Hôpital Régional de Nzérékoré, Nzérékoré, Guinea. ²¹Agence Nationale de Sécurité Sanitaire et Conakry, Guinea. ²²Direction Nationale des Laboratoires, Ministère de la Santé, Conakry, Guinea. ²³Université Gamal Abdel Nasser de Conakry, Conakry, Guinea. ²⁴Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium. ²⁵Infectious Diseases Department, University Hospital Montpellier, Montpellier, France. ²⁶These authors contributed equally: Alpha K. Keita, Fara R. Koundouno, Martin Faye, Ariane Düx, Julia Hinzmann. ²⁷These authors jointly supervised this work: Sophie Duraffour, Ousmane Faye, Fabian Leendertz, Martine Peeters, Abdoulaye Toure, N'Faly Magassouba. ✉e-mail: alpha-kabinet.keita@umontpellier.fr

hospital in Nzérékoré. On 13 February, both of these patients were confirmed to have EVD by the laboratory in Guéckédou, which used a commercial real-time polymerase chain reaction with reverse transcription (qRT–PCR) assay (RealStar Filovirus Screen Kit, Altona Diagnostics). On 13 February, the husband of the index case—who travelled more than 700 km from Gouéké to Conakry, the capital city of Guinea, for treatment—was admitted to the Centre de Traitement Epidémiologique (CTEpi) in Nongo, Ratoma Commune. He presented with fever, nausea, asthenia, abdominal pain and lumbar pain and was strongly suspected to have EVD. A blood sample was analysed on the same day and was found to be positive for Ebola Zaire (*Zaire ebolavirus*; EBOV) according to the GeneXpert molecular diagnostic platform (Xpert Ebola test, Cepheid) and by an in-house qRT–PCR assay. Laboratory confirmation of EVD in the three suspect cases led to the official declaration of the epidemic on 14 February. By 5 March, 14 confirmed cases and 4 probable cases of EVD had been identified, leading to 9 deaths—including 5 confirmed cases as reported by the Agence Nationale de la Sécurité Sanitaire (ANSS) of Guinea. After a period of 25 days without new cases, two new cases were reported around Nzérékoré on 1 and 3 April, and on 19 June 2021 the outbreak was declared to be over. In total, 16 confirmed cases were reported, among which 12 people died.

Genomic characterization of the virus that caused the 2021 epidemic of EVD in Guinea was of immediate importance to public health. First, because diagnostic tools, therapeutics and vaccines with proven effectiveness in recent EVD outbreaks—such as in Guinea (2013–2016) and in the Equateur and North-Kivu/Ituri provinces of the Democratic Republic of the Congo (DRC) (2018–2020)—have primarily been developed for EBOV^{3–5}. Second, to identify whether the outbreak resulted from a new zoonotic transmission event or from the resurgence of a viral strain that had circulated in a previous EBOV outbreak: it is known that EBOV can persist in the bodily fluids of patients who have survived EVD and can be at the origin of new transmission chains^{6–8}. Although the Xpert Ebola test was developed to detect only EBOV strains and the in-house qRT–PCR assay uses a probe that is specifically designed to detect EBOV⁹, additional confirmation by sequence analysis was sought by targeting a short fragment in the viral protein 35 region of the sample from the patient who was hospitalized in Conakry. The phylogenetic tree (Supplementary Fig. 1) underscores that this highly conserved region can discriminate between Ebola virus species, and analysis confirmed that the virus that caused the new outbreak was of the species *Zaire ebolavirus*. This confirmed that available vaccines and the vast majority of molecular-diagnostic tools and therapeutics could be immediately applied.

To gain further insight into the genomic make-up of the viruses causing this outbreak, 11 complete or near-complete (greater than 95% recovery) and 8 partial (greater than 65% recovery) genomic sequences from 12 of the 14 confirmed cases were obtained by 3 different laboratories using different next-generation sequencing technologies (Table 1). To facilitate the public-health response and the evaluation of existing medical countermeasures, sequencing results were made publicly available on 12 March through joint posting (<https://virological.org/c/ebolavirus/guinea-2021/44>). Blood and swab samples from 14 patients with confirmed EVD, sampled from 12 February to 4 March, were processed by the following methods: hybridization capture technology and sequencing on Illumina iSeq100, an amplicon-based protocol with EBOV-specific primer pools and sequencing on MinION (Oxford Nanopore Technologies (ONT)), and a hybrid-capture-based approach using a probe panel that included EBOV-specific targets followed by TruSeq exome enrichment, as previously described⁵. The data generated between the three groups were pooled and the sequence that had the highest quality was chosen for each patient. This enabled us to reconstruct 12 high-quality EBOV genomes that covered 82.9–99.9% of the reference genome (KR534588) (Table 1). The consensus EBOV sequences with the highest genome recovery (greater than 82.9%) from 12 different patients were used in further analyses.

Maximum likelihood phylogenetic reconstruction places the 12 genomes from the 2021 outbreak of EVD in Guinea as a single cluster among the EBOV viruses that were responsible for the 2013–2016 outbreak in West Africa (Figs. 1, 2). The genomes from the 2021 outbreak share 10 substitutions (compared with KJ660346) that were accumulated during the 2013–2016 outbreak, including the A82V marker mutation for human adaptation in the glycoprotein that arose when the virus spread to Sierra Leone^{11,12}. These patterns provide strong evidence of a direct link to human cases from the 2013–2016 outbreak rather than a new spillover from an animal reservoir. The 2021 lineage is nested within a clade that predominantly consists of genomes sampled from Guinea in 2014 (Fig. 2). The branch by which the 2021 cluster diverges from the previous outbreak exhibits only 12 substitutions, which is far fewer than would be expected from the evolution of EBOV during 6 years of sustained human-to-human transmission (Fig. 3). Using a local molecular-clock analysis, we estimate a 6.4-fold (95% highest posterior density (HPD) interval: 3.3-fold, 10.1-fold) lower rate along this branch. For comparison, we also estimate a 5.5-fold (1.6-fold, 10.8-fold) lower rate along the branch leading to the 2016 cases, which were linked to a patient who survived the disease and in whom the virus persisted for more than 500 days^{7,13}. Rather than a constant long-term low evolutionary rate, some degree of latency or dormancy during persistent infection seems to be a more likely explanation for the low divergence of the genomes from the 2021 epidemic. We tested whether the 12 genomes from the 2021 epidemic, which were sampled over a time period of less than one month, contained sufficient temporal signal to estimate the time to most recent common ancestor (tMRCA) (Supplementary Fig. 2); however, we did not identify statistical support for sufficient divergence accumulation over this short timescale. We therefore calibrated our analysis using an evolutionary rate that reflects EBOV evolution under sustained human-to-human transmission (as estimated by the local molecular-clock analysis). This resulted in a tMRCA estimate of 22 January 2021 (95% HPD interval: 29 December 2020, 10 February 2021).

These results open up a new perspective on the relatively rare observation of EBOV re-emergence. It is assumed that all known filovirus outbreaks in humans are the result of independent zoonotic transmission events from bat reservoir species or from intermediate or amplifying hosts such as apes and duikers⁶. Here we clearly show that, even almost five years after the declaration of the end of an epidemic, new outbreaks could also be the result of transmission from humans who were infected during a previous epidemic. The viruses from the 2021 outbreak fall within the lineage of EBOV viruses obtained from humans during the 2014–2016 outbreak; as such, it is very unlikely that this new outbreak has an animal origin or is the result of a new cross-species transmission with the same lineage that remained latent in this natural host, which in that scenario would be at the basis of the West African cluster. The limited genomic divergence between 2014–2015 and 2021 is compatible with a slow long-term evolutionary rate. However, a relatively long phase of latency might be more likely than continuous slow replication. Independent of the mechanistic explanation, the virus most probably persisted at a low level in a human who had survived previous infection. Plausible scenarios of EBOV transmission to the index case include: sexual transmission by exposure to EBOV in semen from a male survivor; contact with body fluids from a survivor who had a relapse of symptomatic EVD (for example during healthcare—the index case was a healthcare worker); or relapse of EVD in the index case—although she was not known to have been infected previously, she could have had an asymptomatic or pauci-symptomatic EBOV infection during the previous outbreak. A detailed investigation of the family of the index case by anthropologists revealed that she was not known to have had EVD previously, nor were her husband or close relatives. However, among more distantly related family, 25 individuals had EVD during the previous outbreak. Only five survived, although the index case apparently had no recent contacts with this part of the family. Consultation of the hospital registers in Gouéké

Table 1 | Patient and sample characteristics and sequencing results obtained by the laboratories involved in the study

Patient	Sex	Age (years)	Date of sampling	CERFIG		PFHG		IPD	
				C _t value ^a	Genome recovery (%)	C _t value ^a	Genome recovery (%)	C _t value ^a	Genome recovery (%)
1	F	54	12 February 2021 ^b	–	–	22.4	87.8	29.3	99.6
			19 February 2021 ^c	33.1	0.3	–	–	–	–
2	F	70	12 February 2021 ^b	–	–	25.9	67.8	37.1	98.7
3	M	61	13 February 2021 ^b	29.4	5.0	Neg.	7.5	Neg.	2.5
4	M	46	20 February 2021 ^c	24.3	12.6	–	–	–	–
5	M	22	22 February 2021 ^b	32.5	99.4	23.2	93.3	–	–
6	M	65	23 February 2021 ^c	–	–	20.5	97.3	–	–
7	F	75	26 February 2021 ^b	–	–	19.5	95.5	–	–
8	M	29	26 February 2021 ^b	–	–	18.8	98.1	–	–
9	M	32	26 February 2021 ^b	24.6	99.9	27.8	77.6	–	–
10	F	30	26 February 2021 ^b	26.0	99.7	23.0	82.2	–	–
11	F	55	26 February 2021 ^b	36.4	75.4	28.8	82.9	–	–
12	M	45	26 February 2021 ^b	–	–	20.5	96.1	–	–
13	M	42	26 February 2021 ^b	25.0	99.9	29.3	70.4	–	–
14	F	40	4 March 2021 ^b	–	–	22.0	97.5	–	–

^aCycle threshold (C_t) value was measured in the sequencing laboratory before starting the sequencing process.

^bPatient samples obtained at diagnosis.

^cFollow-up samples from patients.

All samples from patients were whole EDTA blood, except for patient 12 for whom a swab was used.

– indicates that the sample was not tested in this laboratory. The values differ between laboratories owing to possible degradation of the sample or the RNA during transport and storage.

Neg., negative.

showed that all patients seen by the index case in January 2021 were in good health and were still in good health in March 2021. However, the index case also performed informal consultations outside the hospital environment, which could not be verified. An alternative scenario is that the nurse was not the actual index case, but was part of a small, unrecognized chain of human-to-human transmission in this area of Guinea. However, the diversity of the currently available genomes is limited, and molecular-clock analysis suggests a recent tMRCA, with a mean estimate close to the time that the nurse was first hospitalized and a 95% HPD boundary around the beginning of the year. This provides some reassurance that the outbreak was detected early.

The 2013–2016 outbreak in West Africa was the largest and most complex recent outbreak of EBOV, and involved more than 28,000 cases,

11,000 deaths and an estimated 17,000 survivors, mostly in Guinea, Liberia and Sierra Leone². This large outbreak provided new information about the disease itself as well as about the medical, social and psychological implications for patients who survived the disease^{14–16}. It was also possible to estimate, to some extent, the proportions of asymptomatic or pauci-symptomatic infections and to identify their role in specific unusual transmission chains^{17–19}. Although the main route of human-to-human transmission of EBOV is direct contact with infected bodily fluids from symptomatic or deceased patients, some transmission chains in this outbreak were associated with viral persistence in semen³. Several studies demonstrated viral persistence in more than 50% of male survivors at 6 months after discharge from Ebola treatment units (ETU), and the maximum duration of persistence in semen has been reported to be up to 500–700 days after ETU discharge in a small number of male EVD survivors^{9,20–22}. Transmission through other bodily fluids (such as breast milk and cervicovaginal fluids) is also suspected^{18,23–25}. Furthermore, some immunological studies among survivors suggest a continuous or intermittent EBOV antigenic stimulation due to persistence of an EBOV reservoir in some survivors^{26,27}, although this was not confirmed in another study²⁸. Cases of relapse of EVD have also been sporadically reported and could be the origin of large transmission chains, as recently reported in the North-Kivu outbreak in DRC²⁹. For example, the presence of EBOV RNA, 500 days after ETU discharge, in the breast milk of a woman who was not pregnant when she developed EVD has recently been reported. She attended the hospital owing to complications at 8 months of pregnancy, and a breast milk sample that was taken 1 month after delivery tested positive for EBOV RNA⁹. These examples illustrate that healthcare workers can be exposed to EBOV when taking care of patients who survived EVD but have an unrecognized relapse of their infection. The 2021 outbreak now highlights that viral persistence and reactivation is not limited to a two-year period, but can also occur on much longer timescales with late reactivation.

Active genomic surveillance has already shown the resurgence of previous strains in other outbreaks of the disease. For example, two

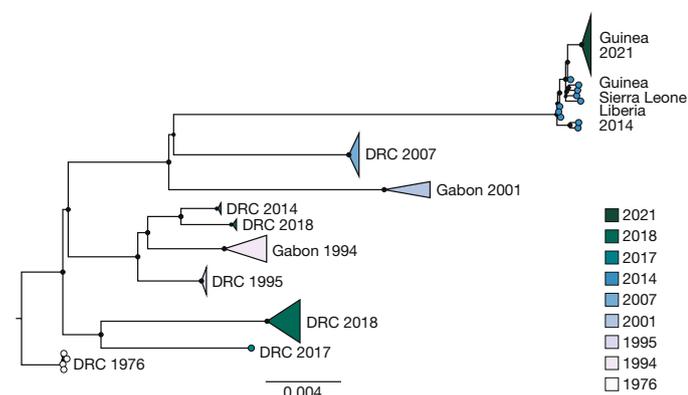


Fig. 1 | Maximum likelihood phylogenetic reconstruction for 55 representative genomes from previous outbreaks of Zaire ebolavirus and 12 genomes from the 2021 outbreak in Guinea. Most clades for single or multiple closely related outbreaks are collapsed and internal node support is proportional to the size of the internal node circles. The clades or tip circles are labelled with the locations and years of the outbreaks, and coloured according to the (first) year of detection.

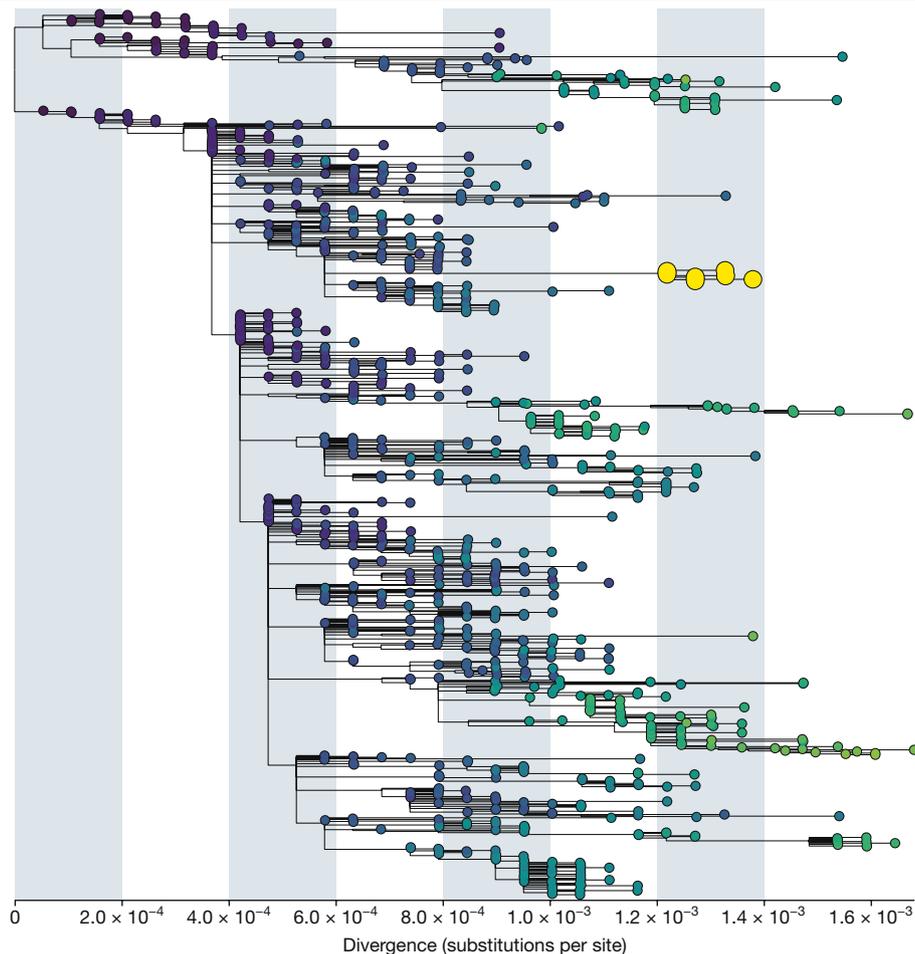


Fig. 2 | Maximum likelihood phylogenetic reconstruction for 1,065 genomes sampled during the 2013–2016 West African outbreak and 12 genomes from the 2021 outbreak in Guinea. A colour gradient (from purple

to green for increasing divergence) is used to colour the tip circles. The 2021 genomes are shown with a larger circle in yellow.

EBOV variants circulated simultaneously within the same region during the recent 2020 outbreak in Equateur province, DRC³⁰. Moreover, strains from the two consecutive outbreaks in Luebo, DRC, in 2007 and 2008, are also so closely related that it now seems difficult to exclude that the epidemic observed in 2008 was due to a resurgence event from

patient who survived EVD in the 2007 outbreak^{31,32}. However, the limited genomic sampling does not allow for a formal test of this hypothesis.

Although the majority of EVD outbreaks remained limited both in the number of cases and in geographic spread, the two largest outbreaks in West Africa (December 2013–June 2016) and in eastern DRC (August

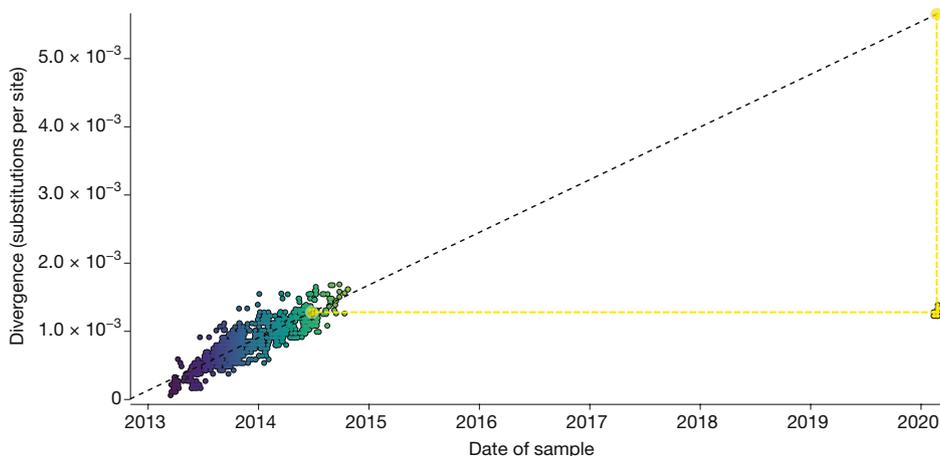


Fig. 3 | Temporal divergence plot showing genetic divergence from the root over time. This plot relates to the tree shown in Fig. 2. The regression is exclusively fitted to genomes sampled between 2014 and 2015. The same colours are used for the data points as in Fig. 2. The dashed yellow lines

highlight how the 2021 data points deviate from the relationship between sampling time and sequence divergence. According to this relationship, about 95 substitutions (95% prediction interval: 88–101) are expected on the branch ancestral to the 2021 cluster, whereas only 12 are inferred on this branch.

2018–June 2020) infected thousands of individuals over wide geographic areas, leading to large numbers of EVD survivors. This means that the risk of resurgence is higher than ever before. Continued surveillance of EVD survivors is therefore warranted to monitor the reactivation and relapse of EVD infection and the potential presence of the virus in bodily fluids. This work and associated communications must be conducted with the utmost care for the wellbeing of EVD survivors. During the 2013–2016 outbreak in Guinea, patients who survived EVD had a mixed experience after discharge from ETUs. On the one hand, they were considered as heroes by non-governmental organizations and became living testimonies of a possible recovery^{33,34}. On the other hand, they experienced different forms of stigmatization, such as rejection by family and friends, refusal of involvement in collective work, loss of jobs and housing, and sometimes self-isolation from social life and workplaces³⁵. The human origin of the 2021 EVD outbreak, and the associated shift in our perception of EBOV emergence, call for careful attention to survivors of the disease. The concern that survivors will be stigmatized as a source of danger should be a matter of scrupulous attention³⁶. This is especially true for the area of Gouécké, which is only 9 km away from Womey—a village that is emblematic of the violent reaction of the population towards the EVD response team during the 2013–2016 epidemic³⁷.

Since the 2013–2016 EVD outbreak in Western Africa, genome sequencing has become a major component of the response to outbreaks^{10,38–41}. The establishment of in-country sequencing and the building of capacity enabled a timely characterization of EBOV strains in the 2021 outbreak in Guinea. In addition to the importance of appropriate healthcare measures focused on survivors, the late resurgence of the virus also highlights the urgent need for further research into potent antiviral agents that can eradicate the latent virus reservoir in patients with EVD, and into efficient vaccines that provide long-term protection. In parallel, vaccination could also be considered to boost protective antibody responses in survivors of the disease²⁷. The vaccination of populations in areas with previous EBOV outbreaks could also be promoted to prevent secondary cases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03901-9>.

- Baize, S. et al. Emergence of Zaire Ebola virus disease in Guinea. *N. Engl. J. Med.* **371**, 1418–1425 (2014).
- World Health Organization. *2016 Situation Report: Ebola Virus Disease* (World Health Organization, 2016).
- Malvy, D., McElroy, A. K., de Clerck, H., Günther, S. & van Griensven, J. Ebola virus disease. *Lancet* **393**, 936–948 (2019).
- Mbala-Kingebeni, P. et al. Rapid confirmation of the Zaire Ebola virus in the outbreak of the Equateur Province in the Democratic Republic of Congo: implications for public health interventions. *Clin. Infect. Dis.* **68**, 330–333 (2019).
- Mbala-Kingebeni, P. et al. 2018 Ebola virus disease outbreak in Equateur Province, Democratic Republic of the Congo: a retrospective genomic characterization. *Lancet Infect. Dis.* **19**, 641–647 (2019).
- Pigott, D. M. et al. Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**, e04395 (2014).
- Diallo, B. et al. Resurgence of Ebola virus disease in Guinea linked to a survivor with virus persistence in seminal fluid for more than 500 days. *Clin. Infect. Dis.* **63**, 1353–1356 (2016).
- Sissoko, D. et al. Ebola virus persistence in breast milk after no reported illness: a likely source of virus transmission from mother to child. *Clin. Infect. Dis.* **64**, 513–516 (2017).
- Keita, A. K. et al. A 40-month follow-up of Ebola virus disease survivors in Guinea (PostEboGui) reveals long-term detection of Ebola viral ribonucleic acid in semen and breast milk. *Open Forum Infect. Dis.* **6**, ofz482 (2019).
- Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
- Diehl, W. E. et al. Ebola virus glycoprotein with increased infectivity dominated the 2013–2016 epidemic. *Cell* **167**, 1088–1098.e6 (2016).
- Urbanowicz, R. A. et al. Human adaptation of Ebola virus during the West African outbreak. *Cell* **167**, 1079–1087.e5 (2016).
- Blackley, D. J. et al. Reduced evolutionary rate in reemerged Ebola virus transmission chains. *Sci. Adv.* **2**, e1600378 (2016).
- Diallo, M. S. K. et al. Understanding the long-term evolution and predictors of sequelae of Ebola virus disease survivors in Guinea: a 48-month prospective, longitudinal cohort study (PostEboGui). *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab168> (2021).
- Etard, J. F. et al. Multidisciplinary assessment of post-Ebola sequelae in Guinea (PostEboGui): an observational cohort study. *Lancet Infect. Dis.* **17**, 545–552 (2017).
- PREVAIL III Study Group. A longitudinal study of Ebola sequelae in Liberia. *N. Engl. J. Med.* **380**, 924–934 (2019).
- Diallo, M. S. K. et al. Prevalence of infection among asymptomatic and paucisymptomatic contact persons exposed to Ebola virus in Guinea: a retrospective, cross-sectional observational study. *Lancet Infect. Dis.* **19**, 308–316 (2019).
- Glynn, J. R. et al. Asymptomatic infection and unrecognised Ebola virus disease in Ebola-affected households in Sierra Leone: a cross-sectional study using a new non-invasive assay for antibodies to Ebola virus. *Lancet Infect. Dis.* **17**, 645–653 (2017).
- Camara, I. et al. Unrecognized Ebola virus infection in Guinea: complexity of surveillance in a health crisis situation: case report. *Pan Afr. Med. J.* **36**, 201 (2020).
- Thorsen, A. E. et al. Persistence of Ebola virus in semen among Ebola virus disease survivors in Sierra Leone: a cohort study of frequency, duration, and risk factors. *PLoS Med.* **18**, e1003273 (2021).
- Fischer, W. A. et al. Ebola virus ribonucleic acid detection in semen more than two years after resolution of acute Ebola virus infection. *Open Forum Infect. Dis.* **4**, ofx155 (2017).
- Sissoko, D. et al. Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study. *Lancet Glob. Health* **5**, e80–e88 (2017).
- Mate, S. E. et al. Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* **373**, 2448–2454 (2015).
- Dokubo, E. K. et al. Persistence of Ebola virus after the end of widespread transmission in Liberia: an outbreak report. *Lancet Infect. Dis.* **18**, 1015–1024 (2018).
- Liu, W. J. et al. Comprehensive clinical and laboratory follow-up of a female patient with Ebola virus disease: Sierra Leone Ebola virus persistence study. *Open Forum Infect. Dis.* **6**, ofz068 (2019).
- Wiedemann, A. et al. Long-lasting severe immune dysfunction in Ebola virus disease survivors. *Nat. Commun.* **11**, 3730 (2020).
- Adaken, C. et al. Ebola virus antibody decay—stimulation in a high proportion of survivors. *Nature* **590**, 468–472 (2021).
- Thom, R. et al. Longitudinal antibody and T cell responses in Ebola virus disease survivors and contacts: an observational cohort study. *Lancet Infect. Dis.* **21**, 507–516 (2021).
- Mbala-Kingebeni, P. et al. Ebola virus transmission initiated by relapse of systemic Ebola virus disease. *N. Engl. J. Med.* **384**, 1240–1247 (2021).
- Muyembe-Tamfum, J. J. et al. Two Ebola virus variants circulating during the 2020 Equateur Province outbreak. <https://virological.org/t/two-ebola-virus-variants-circulating-during-the-2020-equateur-province-outbreak/538> (2020).
- Leroy, E. M. et al. Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Dis.* **9**, 723–728 (2009).
- Grard, G. et al. Emergence of divergent Zaire Ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. *J. Infect. Dis.* **204** (Suppl 3), S776–S784 (2011).
- Enria, L. What Crisis Produces: Dangerous Bodies, Ebola Heroes and Resistance in Sierra Leone' *BathPapers in International Development and Wellbeing*, Centre for Development Studies, University of Bath no. 53 (2017).
- Ronse, M. et al. What motivates Ebola survivors to donate plasma during an emergency clinical trial? The case of Ebola-Tx in Guinea. *PLoS Negl. Trop. Dis.* **12**, e0006885 (2018).
- Sow, S., Desclaux, A., Taverne, B. & Groupe d'étude PostEboGui. Ebola en Guinée: formes de la stigmatisation des acteurs de sante survivants. *Bull. Soc. Pathol. Exot.* **109**, 309–313 (2016).
- Farmer, P. *Aids and Accusation. Haiti and the Geography of Blame* (Univ. California Press, 2006).
- Wilkinson, A. & Fairhead, J. Comparison of social resistance to Ebola response in Sierra Leone and Guinea suggests explanations lie in political configurations not culture. *Crit. Public Health* **27**, 14–27 (2017).
- Gire, S. K. et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
- Arias, A. et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol.* **2**, vew016 (2016).
- Pini, A. et al. Field investigation with real-time virus genetic characterisation support of a cluster of Ebola virus disease cases in Dubréka, Guinea, April to June 2015. *Euro Surveill.* **23**, 17-00140 (2018).
- Mbala-Kingebeni, P. et al. Medical countermeasures during the 2018 Ebola virus disease outbreak in the North Kivu and Ituri Provinces of the Democratic Republic of the Congo: a rapid genomic assessment. *Lancet Infect. Dis.* **19**, 648–657 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Ethics statement

Diagnostic specimens were collected as part of the emergency response from the Ministry of Health of Guinea, and therefore consent for sample collection was waived. All preparation of samples for sequencing, genomic analysis and data analysis was performed on anonymized samples identifiable only by their laboratory or epidemiological identifier.

Confirmation of Ebola virus species by sequence analysis of the VP35 fragment at CERFIG

Viral RNA was extracted from 140 µl of whole blood collected from samples from the patient hospitalized in Conakry, using the Nuclisens kit (Biomerieux) and following the manufacturer's instructions. Amplification of a small fragment of the VP35 region was attempted in a semi-nested PCR with a modified protocol as previously described⁴. First-round VP35 PCR products from positive samples were barcoded and pooled using the Native Barcoding Kit EXP-NBD104 (ONT). Sequencing libraries were generated from the barcoded products using the Genomic DNA Sequencing Kit SQK-LSK109 (ONT) and were loaded onto a R9 flow cell on a MinION (ONT). Genetic data were collected for 1 h. Basecalling, adapter removal and demultiplexing of .fastq files were performed with MinKNOW, v.4.1.22. Fastq reads >Q11 were used for mapping a virus database with the Genome Detective tool (<https://www.genomedetective.com/app/typingtool/virus/>). The generated consensus sequence was used for further analysis. For phylogenetic inference, we retrieved one sequence per outbreak from the haemorrhagic fever virus (HFV) database to which we added the newly generated VP35 sequence of the new outbreak. Phylogenetic analyses were performed using maximum likelihood methods using IQ-TREE with 1,000 bootstraps for branch support^{42,43}. The general time-reversible (GTR) model plus a discrete gamma distribution were used as nucleotide substitution models.

Full-length genome sequencing of the new Ebola viruses

Genome sequencing at CERFIG. Whole-genome sequencing was attempted on viral extracts for samples that were positive for EBOV glycoprotein (GP) and nucleoprotein (NP) on the GeneXpert molecular diagnostic platform (Xpert Ebola Assay) with the GP and NP of *Zaire ebolavirus*. We extracted full nucleic acid using the QIAamp Viral RNA Mini Kit (Qiagen). After DNase treatment with TURBO DNA-free Kit (Ambion) and clean-up with RNA Clean & Concentrator Kit (Zymo Research), RNA was converted to double-stranded cDNA (ds-cDNA) using the SuperScript IV First-Strand Synthesis System (Invitrogen) and NEBNext mRNA Second Strand Synthesis Module (New England Biolabs). The resulting ds-cDNA was enzymatically fragmented with NEBNext dsDNA Fragmentase (New England Biolabs) and converted to dual indexed libraries with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) and NEBNext Multiplex Oligos for Illumina (New England Biolabs). To enrich EBOV in the libraries, we performed two rounds of hybridization capture (16 h at 65 °C) with custom-made biotinylated RNA baits (120 nucleotides, 2-fold tiling; Arbor Biosciences) covering representative genomes for *Zaire ebolavirus* (KC242801), *Sudan ebolavirus* (KC242783), *Reston ebolavirus* (NC_004161), *Tai Forest ebolavirus* (NC_014372), *Bundibugyo ebolavirus* (KC545395) and *Marburg marburgvirus* (FJ750956), following the myBaits Hybridization Capture for Targeted NGS protocol (v.4.01). After the second round, capture products were quantified using the Qubit 3.0 Fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen), and pooled in equimolar amounts for sequencing on an Illumina iSeq using iSeq 100 i1 Reagents (2 × 150 cycles). Sequencing reads were filtered (adapter removal and quality filtering) with Trimmomatic⁴⁴ (settings: LEADING:30 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:40), merged with ClipAndMerge (<https://github.com/apeltzer/ClipAndMerge>), and mapped to the *Zaire ebolavirus* RefSeq genome (NC_002549) using

BWA-MEM⁴⁵. Mapped reads were sorted and deduplicated with SortSam and MarkDuplicates from the Picard suite (Broad Institute, Picard; <http://broadinstitute.github.io/picard>). We generated consensus sequences using Geneious Prime 2020.2.3 (<https://www.geneious.com>), in which unambiguous bases were called when at least 90% of at least 20 unique reads were in agreement (20×, 90%). For samples with few mapped reads (0001, 0002, 0010, 0030), we also called a consensus at 2×, 90% and 5×, 90%.

Genome sequencing at PFHG. Sequencing at PFHG was performed using a mobile MinION facility deployed by BNITM to Guinea at the beginning of March 2021. A total of 13 EBOV-positive initial diagnostic samples processed at the Laboratoire des Fièvres Hémorragiques Virales de Gueckédou, the Laboratoire Régional de l'Hôpital de Nzérékoré were used for sequencing. If RNAs from diagnostic procedures performed by the peripheral laboratories was not sent to PFHG, samples were inactivated and RNA was extracted from 50 µl for whole blood EDTA, 70 µl of plasma from EDTA blood or from 140 µl of wet swabs using the QIAamp Viral RNA Mini Kit (Qiagen) following the manufacturer's instructions. Tiled primers generating overlapping products combined with a highly multiplexed PCR protocol were used for amplicon generation¹⁰. At start of deployment, three different primer pools (V3 or pan_10_EBOV, V4 or pan_EBOV and Zaire-PHE or EBOV-Zaire-PHE) were tested and results were combined for the optimal recovery of consensus. A new primer pool V5 (EBOV-Makona-V5) was further designed and implemented to increase consensus recovery. Primer pools V3, V4 and V5 were designed by the ARTIC network and Zaire-PHE primer pools by Public Health England (PHE). For V3, 62 primers were used, while for V4 and V5, 61 primers pairs were used, to amplify products of around 400 nt in length. For Zaire-PHE, 71 primer pairs were used to amplify products of around 350 nt in length for the approximately 20 kb viral genome. All primer pools used can be found in Supplementary Table 1. The multiplex PCR was performed as described by the most up-to-date ARTIC protocol for nCoV-2019 amplicon sequencing (nCoV-2019 sequencing protocol V3 (LoCost) V.3 (<https://artic.network/ncov-2019>), adapted to include the EBOV-specific primer sets. In brief, RNA was directly used for cDNA synthesis using the LunaScript RT SuperMix (New England Biolabs) and the cDNA generated was used as template in the multiplex PCR, which was performed in two reaction pools using Q5 Hot Start DNA Polymerase (New England Biolabs). The resulting amplicons from the two PCR pools were pooled in equal volumes and the pooled amplicons were diluted 1:10 with nuclease-free water.

Sequencing libraries were prepared, barcoded and multiplexed using the Ligation Sequencing Kit (SQK-LSK109) from ONT combined with the Native Expansion pack (EXP-NDB104, EXP-NBD114, EXP-NBD196) following the ARTIC Network's library preparation protocol (nCoV-2019 sequencing protocol v3 (LoCost) V.3 (<https://artic.network/ncov-2019>)). For the preparation of fewer than 11 samples, each sample was prepared in multiples to achieve the library concentration required for sequencing. In brief, the diluted pooled amplicons were end-repaired using the Ultra II End Prep Module (New England Biolabs) followed by barcode ligation using the Blunt/TA Ligase Master Mix and one unique barcode per sample. Equal volumes from each native barcoding reaction were pooled and subject to bead clean-up using 0.4× AMPure beads. The pooled barcoded amplicons were quantified using the Qubit Fluorometer (Thermo Fisher Scientific) and AMII adapter ligation was performed using the Quick T4 DNA Ligase (New England Biolabs) followed by an additional bead clean-up. The adaptor-ligated barcoded amplicon pool was quantified using the Qubit Fluorometer (Thermo Fisher Scientific) aiming for a minimum recovery of 15 ng sequencing library to load onto the flow cell.

Sequencing libraries were sequenced using R9.4.1 Flow Cells (FLO-MIN106D, ONT) on the Mk1C device (ONT) using MinKNOW v.21.02.2 with real-time high accuracy base-calling and stringent demultiplexing (minimum barcoding score = 60). Within the barcoding

options, barcoding on both ends and mid-read barcodes were both switched on. Reads were demultiplexed and binned in a barcode specific folder only if a barcode above the minimum barcoding score was identified on both read ends and if mid-read barcodes were not identified. Sequencing runs were stopped after around 24 h, and base-calling was allowed to finish before data handling.

Bioinformatics data analysis was performed as per the ARTIC protocol using a combination of the ARTIC EBOV (<https://artic.network/ebov/ebov-bioinformatics-sop.html>) and ARTIC SARS-CoV-2 (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>) pipelines. A few minor modifications to the ARTIC bioinformatics protocol were incorporated. The two initial steps described, base-calling with Guppy and demultiplexing, were omitted as these were both done on the Mk1C device in real-time during the sequencing run; subsequently, the bioinformatics analysis was initiated from the read-filtering step (ARTIC Guppyplex). In brief, the ARTIC Guppyplex program was used to collect reads for each barcode into a single fastq file, in the presence of a length filter to remove chimeric reads. Reads were filtered based on length with a minimum (option: --min-length) and maximum (option: --max-length) length cut-off based on the amplicon size used (For V3, V4 and V5 primer pools: --min-length 400 and --max-length 700, for Zaire-PHE primer pool: --min-length 350 and --max-length 650). The quality check was omitted because only reads with a quality score of greater than 7 were processed. After merging and filtering, the ARTIC MinION pipeline was used to obtain the consensus sequences. The data were normalized to 200 and, using the --scheme-directory option, the pipeline was directed to the respective primer scheme used for each barcode. Reads were aligned to the NCBI reference KJ660347 (*Zaire ebolavirus* isolate H.sapiens-wt/GIN/2014/Makona-Gueckedou-C07) for data generated using V3, V4, and V5 primer pools and to NC_002549.1 (*Zaire ebolavirus* isolate Ebola virus/H.sapiens-tc/COD/1976/Yambuku-Mayinga) for data generated using Zaire-PHE primer pools.

Sequencing at IPD. Viral RNA was extracted from 140 µl of whole blood samples using the QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer's instructions and eluted in nuclease-free water to a final volume of 60 µl. Extracted RNA was tested using qRT-PCR as previously described⁴⁶. In brief, the DNA library was prepared and enriched using the Illumina RNA Prep with Enrichment (L) Tagmentation kit (Illumina) according to the manufacturer's recommendations with a pan viral probe panel that included EBOV-specific targets⁵. The purified libraries were pooled and sequenced on the Illumina MiSeq platform using the MiSeq Reagents Kit v3 (Illumina) according to the manufacturer's instructions. Illumina sequence reads were quality trimmed by Prinseq-lite and consensus EBOV genome sequences were generated using an in-house de novo genome assembly pipeline.

Phylogenetic analysis of full-length genome sequences

Phylogenetic inference. The new EBOV genome sequences were embedded in different datasets for subsequent analyses. For phylogenetic reconstruction, we use a Zaire Ebola virus dataset consisting of 55 representative genomes from previous outbreaks and a Makona virus dataset consisting of 1,065 genomes sampled from Guinea, Sierra Leone and Liberia between 2014 and 2015. Multiple sequence alignment was performed using mafft⁴⁷. We identified 6 T-to-C mutations in the genome from patient 11 that were indicative of mutations induced by adenosine deaminases acting on RNA. According to previous recommendations⁴⁸, we masked these positions in this genome in all further analyses. Maximum likelihood trees were reconstructed using IQ-TREE under the GTR model with gamma (G) distributed rate variation among sites⁴⁹. Temporal divergence plots of genetic divergence from the root of phylogenies against sampling time were constructed using TempEst⁵⁰. To construct the temporal divergence plot for the Guinean 2021 genome data, we used a tree reconstructed under an HKY+G model.

Local molecular-clock model analysis. We used BEAST to fit a local molecular-clock model to a dataset consisting of 1,020 dated Makona virus genomes and one of the 2021 genomes (patient 1)^{51,52}. We specified a separate rate on the tip branch for this genome as well as on the tip branch for a genome in a 2016 outbreak. We used the Skygrid coalescent model as a flexible nonparametric tree prior and an HKY+G substitution model⁵³.

Guinea 2021 tMRC estimation. Temporal signal was evaluated using the BETS procedure⁵⁴. We estimated a slightly lower log marginal likelihood for a model that uses tip dates (-26,063.6) compared to a model that assumes sequences are sampled at the same time (-26,062.1). These BEAST analyses were performed using an exponential growth model, a strict molecular-clock model and an HKY+G substitution model. We specified a lognormal prior with a mean of 1 and a standard deviation of 5 on the population size and a Laplace prior with a scale of 100 on the growth rate. Default priors were used for all other parameters. For the estimation of divergence time, we used a normal prior on the substitution rate with a mean of 0.001 and a standard deviation of 0.00004 based on the background EBOV rate estimated by the local molecular-clock analysis.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequencing results were made publicly available on 12 March 2021 through joint posting on <https://virological.org/c/ebolavirus/guinea-2021/44>. The sequences generated at CERFIG have been deposited to GitHub under project link https://github.com/kabinet1980/Ebov_Guinea2021/blob/main/EBOV_Guinea_2021_genomes_CERFIG.fasta and at the European Nucleotide Archive (ENA) under accession code PRJEB43650. The sequences generated at PFHVG have been deposited to GitHub under project link <https://github.com/PFHVG/EBOVsequencing> and the genome sequences for the two samples at IPD are available at <https://drive.google.com/drive/folders/14fdGdNjWw17TkjrEQKLCrwlJ4WBBH16K>. Genome sequences are also available at the NCBI GenBank under accession codes ERX5245591 to ERX5245598; MZ424849 to MZ424862; MZ605320 and MZ605321.

Code availability

All the codes for the analyses presented in this paper, including the analysis pipeline, is described in detail in Methods and is available in published papers and public websites or, for in-house pipelines, is available upon reasonable request from the corresponding author.

- Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
- Bolger A. M., Lohse M. & Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Weidmann, M., Mühlberger, E. & Hufert, F. T. Rapid detection protocol for filoviruses. *J. Clin. Virol.* **30**, 94–99 (2004).
- Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
- Dudas, G. et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315 (2017).
- Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

Article

51. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
52. Worobey, M., Han, G. Z. & Rambaut, A. A. Synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* **508**, 254–257 (2014).
53. Gill, M. S. et al. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
54. Duchene, S., et al. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* **6**, veaa061 (2020).

Acknowledgements We thank the ANSS and the Ministry of Health of the Republic of Guinea, the healthcare workers (medical doctors, nurses and laboratory technicians) from the treatment centres in Nzérékoré and Nongo, Conakry; the laboratory personnel in Guékédou, INSP (Conakry) and CERFIG (Conakry). CERFIG also acknowledges J. Gogarten and bioinformatics support from RKI. The UK Health Security Agency would like to thank Oxford Nanopore Technologies for the donation of reagents and equipment to support the setting up of the sequencing capacity at PFHG. BNITM thanks the ARTIC Network (<https://artic.network/>). The work of CERFIG and TransVIHMI was supported in part by grants from the EBO-SURSY Project funded by the European Union, International Mixt Laboratory 'RESPIRE' of IRD (Institut de Recherche pour le Développement), Montpellier Université d'Excellence (EBOHEALTH; I-Site MUSE, ANR-16-IDEX-0006) and Institut National de la Santé et de la Recherche Médicale (INSERM)/REACTing (REsearch and ACTion targeting emerging infectious diseases). The work of the RKI was partly funded by the German Ministry of Health 'Global Protection Program' project TRICE. F.L.M. received funding through the program 'EBOVAC3 Bringing a prophylactic Ebola vaccine to licensure' funded by Innovative Medicine Initiative (grant agreement number 800176) and run by London School of Hygiene and Tropical Medicine and INSERM. The work of PFHG and BNITM was supported by the German Federal Ministry of Health through support of the WHO Collaborating Centre for Arbovirus and Hemorrhagic Fever Viruses at the BNITM (agreement ZMV I1-2517WHO005), and through the Global Health Protection Programme (GHPP, agreements ZMV I1-2517GHP-704 and ZMV I1-2519GHP704), and by the Coalition for Epidemic Preparedness Innovations (CEPI). The work of BNITM and the UK Health Security Agency was further supported by the Research and Innovation Programme of the European Union under Horizon 2020 grant agreement no. 871029-EVA-GLOBAL. The European Mobile Laboratory (EMLab) coordinated by BNITM is a technical partner of the WHO Global Outbreak Alert and Response Network (GOARN) and the deployment of EMLab experts and sequencing capacities to Guinea was coordinated and supported by the GOARN Operational Support

Team at WHO/HQ and the WHO country office in Guinea. The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 725422-ReservoirDOCS). The ARTIC Network receives funding from the Wellcome Trust through project 206298/Z/17/Z. P.L. acknowledges support by the Research Foundation – Flanders ('Fonds voor Wetenschappelijk Onderzoek – Vlaanderen', G066215N, G05117N and G0B9317N).

Author contributions A.A.S., A.K.K., A. Toure, E.D., F.H.L., F.R.K., L.E.K., M. Peeters, N'F.M., O.F., S.D. and S.G. conceived and designed the study. A.B., A. Doré, A.K.K., A.K.S., A.M.-S., A.S., B.S., C.D.S., D.D., F.L., F.L.M. F.R.K., F.Y.S., F.A.T., F.B.S., G.A., H.D., I.C., J.A.B., J.C., K.I., Kaka Kourouma, Karifa Kourouma, S.L.M., M.B., M.B.K., M.C., M.D.B., M.K., M. Povogui, M.S.S., N.V., N.J.V., S.D. and Y.S. collected data and/or performed medical examinations and/or laboratory diagnostics. A.A., A. Düx, A. Renevey, B.S., H.D., J.A.B., J.H., K.I., K.P., M.F., M.M.D. and S.C.-S. performed sequencing and/or sequence validation. A.A., A. Rambaut, J.H., J.T.M., L.E.K., M.R.W., P.L., S.C.-S. and S.D. performed formal phylogenetic analysis. A. Diallo, C.L., F.L.M. and M. Diop performed data analysis. E.D., F.H.L., M. Peeters, M.W.C., S.D. and S.G. acquired funding. F.H.L., K.P., M.R.W. and S.C.-S. provided reagents. A.K.K., A.L., A. Thielebein, A. Toure, C.P., E.D., F.H.L., G.A.K.-Z., J.Q., M. Diakite, M. Pahlmann, M. Peeters, M.S.B., M.W.C., N'F.M., N.T., P.F., S.D., S.G., S.K. and S.T.P. implemented the project. A. Düx, A.K.K., M. Peeters and S.C.-S. wrote the first draft of the manuscript. A.A.S., A. Düx, A.K.K., A.M.-S., A. Renevey, A. Rambaut, A. Toure, E.D., F.H.L., F.L., F.L.M. L.E.K., M.F., M.M.D., M. Peeters, N'F.M., O.F., P.L., S.C.-S., S.D. and S.G. wrote and edited the manuscript. All authors read and approved the contents of the manuscript.

Competing interests M.W.C. received materials for this study from Oxford Nanopore Technologies. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03901-9>.

Correspondence and requests for materials should be addressed to Alpha Kabinet Keita.

Peer review information *Nature* thanks Robert Garry, Joel Montgomery and Michael Worobey for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection we did not use software to collect data.

Data analysis All the code for the analyses presented in this paper, including the analysis pipeline is described in detail in methods and is available in published papers, public websites or in-house pipelines available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequences generated by CERFIG are available at: https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=724447
 With Accession number : ERX5245598, ERX5245597, ERX5245596, ERX5245595, ERX5245594, ERX5245593, ERX5245592, ERX5245591
 Sequences generated by BNI/PFHG are available at:
 (<https://www.ncbi.nlm.nih.gov/nuccore/MZ424849>)
 (<https://www.ncbi.nlm.nih.gov/nuccore/MZ424862>)
 With Accession numbers : MZ424849, MZ424862
 Sequences generated by IPD are available with these accession numbers: MZ605320, MZ605321

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The study is descriptive and observational analyzing viral genomes from an Ebolavirus outbreak on samples from patients diagnosed infected with Ebolavirus, therefore there was no sample size calculation
Data exclusions	samples from the 14 patients diagnosed with Ebola virus have been included to obtain full length viral sequences, samples were excluded for further analysis when less than 80% of genome sequence was obtained.
Replication	Among the 14 patients included, samples from 7 patients have been tested in two different laboratories, and for one patient samples were tested in three different laboratories.
Randomization	Given that this is an observational study during an acute outbreak, no randomization has been done and almost all patients with confirmed laboratory diagnosis were included
Blinding	This was an observational study without interventions and no blinding was done

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	NA
Wild animals	NA
Field-collected samples	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Among the 14 patients studied, 6 were female and 8 were male; age ranged between 22 and 75 years.
Recruitment	This was an observational study during an acute Ebola outbreak and patients were consecutively included when diagnosed as infected with Ebolavirus using molecular diagnostic assays.

Ethics oversight

Diagnostic specimens were collected as part of the emergency response from the Ministry of public health from Guinea, and therefore consent for sample collection was waived. All preparation of samples for sequencing, genomic analysis and data analysis was performed on anonymized samples identifiable only by their laboratory or epidemiological identifier.

Note that full information on the approval of the study protocol must also be provided in the manuscript.