

# UCLA

## UCLA Previously Published Works

### Title

Predicting the evolution of the Lassa virus endemic area and population at risk over the next decades

### Permalink

<https://escholarship.org/uc/item/15f3k74z>

### Journal

Nature Communications, 13(1)

### ISSN

2041-1723

### Authors

Klitting, Raphaëlle  
Kafetzopoulou, Liana E  
Thiery, Wim  
[et al.](#)

### Publication Date

2022

### DOI

10.1038/s41467-022-33112-3

Peer reviewed

# Predicting the evolution of the Lassa virus endemic area and population at risk over the next decades

---

Received: 30 September 2021


---

Accepted: 2 September 2022

---

Published online: 27 September 2022

---

 Check for updates

---

Raphaëlle Klitting<sup>1</sup>✉, Liana E. Kafetzopoulou<sup>2,3</sup>, Wim Thiery<sup>4</sup>, Gytis Dudas<sup>5</sup>, Sophie Gryseels<sup>6,7</sup>, Anjali Kotamarthi<sup>1</sup>, Bram Vrancken<sup>2</sup>, Karthik Gangavarapu<sup>1</sup>, Mambu Momoh<sup>8,9</sup>, John Demby Sandi<sup>9</sup>, Augustine Goba<sup>9</sup>, Foday Alhasan<sup>9</sup>, Donald S. Grant<sup>9,10</sup>, Sylvanus Okogbenin<sup>11,12</sup>, Ephraim Ogbaini-Emovo<sup>11</sup>, Robert F. Garry<sup>13,14,15</sup>, Allison R. Smither<sup>13</sup>, Mark Zeller<sup>1</sup>, Matthias G. Pauthner<sup>1</sup>, Michelle McGraw<sup>1</sup>, Laura D. Hughes<sup>16</sup>, Sophie Duraffour<sup>3,17</sup>, Stephan Günther<sup>3,17</sup>, Marc A. Suchard<sup>18,19,20</sup>, Philippe Lemey<sup>2</sup>, Kristian G. Andersen<sup>1,21</sup> & Simon Dellicour<sup>2,22</sup>✉

---

Lassa fever is a severe viral hemorrhagic fever caused by a zoonotic virus that repeatedly spills over to humans from its rodent reservoirs. It is currently not known how climate and land use changes could affect the endemic area of this virus, currently limited to parts of West Africa. By exploring the environmental data associated with virus occurrence using ecological niche modelling, we show how temperature, precipitation and the presence of pastures determine ecological suitability for virus circulation. Based on projections of climate, land use, and population changes, we find that regions in Central and East Africa will likely become suitable for Lassa virus over the next decades and estimate that the total population living in ecological conditions that are suitable for Lassa virus circulation may drastically increase by 2070. By analysing geotagged viral genomes using spatially-explicit phylogeography and simulating virus dispersal, we find that in the event of Lassa virus being introduced into a new suitable region, its spread might remain spatially limited over the first decades.

Along with other viral infections that have gained prominence in recent years<sup>1–3</sup>, Lassa fever (Lassa) is listed by the World Health Organization (WHO) as one of the diseases that pose the greatest public health risk<sup>4,5</sup>. Lassa is a viral hemorrhagic fever with variable but generally high case fatality rates<sup>6</sup> for which efficacious countermeasures are lacking<sup>7,8</sup>. There is currently no vaccine approved to prevent Lassa. Although several candidates have shown promising results during preclinical studies, only one (INO-4500) has progressed to clinical trials (now in phase IB, NCT03805984)<sup>9,10</sup>. Regarding Lassa treatment, the only antiviral drug available is the nucleoside analog ribavirin<sup>11</sup>, which is often ineffective<sup>7</sup>. To date, Lassa cases have mostly been reported in West Africa including Guinea, Liberia, Nigeria, and Sierra Leone. While these countries seem to constitute most endemic

hotspots<sup>12</sup>, local under-reporting could potentially bias this overall picture. Nigeria, in particular, has witnessed a significant increase in incidence in recent years, and confirmed more than a thousand cases in 2020<sup>13</sup>. Increasingly, neighbouring countries, including Benin, Ghana, Ivory Coast, Mali, and Togo, have also been reporting infections<sup>14–17</sup>, suggesting that the true Lassa range may span a sizable part of West Africa.

Lassa is caused by Lassa virus<sup>18</sup>, a member of the *Arenaviridae* family (genus *Mammarenavirus*). Human infections are generally thought to occur through direct contact or exposure to the excreta of infected *Mastomys natalensis*<sup>19–23</sup>, although the main transmission mechanism remains to be formally established. *M. natalensis* rodents often live in close contact with human communities<sup>24</sup> and are regarded

as the primary reservoir for the virus<sup>25</sup>. Humans likely contribute little to virus transmission and are considered dead-end hosts, based on studies of rodent biology<sup>18</sup>, ecology<sup>11,26–28</sup>, transmission dynamics<sup>22,29</sup>, and viral genomes<sup>30–32</sup>. While the virus can only spread where its reservoir is present, the range of *M. natalensis* extends beyond that of Lassa virus, spanning most of sub-Saharan Africa<sup>33,34</sup>. The factors underlying this difference in range between the virus and its reservoir have led to long-standing questions about suitability and may be multifactorial: Lassa virus might exclusively circulate within one of the six *M. natalensis* phylogroups or subspecies, namely A-1<sup>35</sup>, which is only found in West Africa<sup>35–37</sup>; other viruses present in *M. natalensis* populations may prevent Lassa virus circulation through competition<sup>38,39</sup>; closely related mammarenaviruses inducing cross-reactive immunity<sup>40–42</sup> may prevent Lassa virus infection; and finally, Lassa virus prevalence may be influenced by different environmental determinants than its reservoir, as in the case of Sin Nombre virus<sup>43</sup>. For this other rodent-borne virus, environmental conditions can impact the abundance of the host, driving the population density of the reservoir below the threshold needed for virus maintenance<sup>43–45</sup>.

Ecological niche modelling studies have identified—although not always agreed on—environmental factors that correlate with the occurrence of Lassa virus infections in rodent and human hosts<sup>34,46–49</sup>. Additional biological and socio-ecological factors may further influence spill-over dynamics, as suggested by mechanistic modelling investigations<sup>46,49,50</sup>. Most of these studies have mapped the current risk for Lassa virus infection in West Africa, identifying risk areas across much of this region<sup>34,46–49</sup>. Like the rest of the world, African countries will increasingly be affected by climate change, with warming temperatures and more extreme, yet rarer, precipitation<sup>51–53</sup>. These changes, combined with an increasing pressure on land resources due to a considerable projected human population expansion, are expected to result in important transformations of land use throughout Africa<sup>54–56</sup>. Previously, Redding and colleagues showed that spill-over events would at least double by 2070 within the Lassa-endemic western African region due to climate change, human population growth, and to a smaller extent, land use changes<sup>46</sup>. It is not known, however, how these environmental changes may affect the distribution of the virus itself<sup>57</sup>.

In this work, we add to previous modelling studies by combining ecological niche modelling and phylogeographic analyses to investigate how the endemic range of Lassa virus may evolve in the next five decades in response to climate change, human population growth, and land use changes. We start with using ecological niche modelling to assess whether regions outside the current endemic range may become suitable for Lassa virus due to changes in environmental conditions. To identify factors driving suitability for virus circulation, we analyse the environmental data associated with the occurrence of Lassa virus for a set of putative explanatory factors. We find that annual mean temperature, annual precipitation, and the presence of pastures are the main factors determining ecological suitability for Lassa virus circulation. Using projections of climate, land use, and population up to 2070, we estimate future ecological suitability and show that within decades, the range suitable for Lassa virus may extend well beyond West Africa. Using population projections, we then estimate that the extended part of the suitable range will be home to an increasingly large number of people in the next decades. We hypothesise that if Lassa virus is introduced into a new suitable region, it could in theory spread there because its reservoir host is present. This process, however, might take time. To gain insights into how fast the virus may be able to spread through a suitable environment, we subsequently perform spatially-explicit phylogeographic analyses. Using geotagged viral genomes, we show that over the first decades following a successful introduction into a new region its propagation could remain spatially limited, unless the virus spreads significantly faster than in current endemic areas. By combining ecological niche modelling with

spatially-explicit phylogeography, our study showcases how climate and land use change may transform the future risk of Lassa in Africa.

## Results

### Main determinants of ecological suitability for Lassa virus

To identify factors that determine ecological suitability for Lassa virus and *M. natalensis*, we built ecological niche models, considering temperature, precipitation, seven types of land cover, and human population as potential determinants. Using a boosted regression trees<sup>58</sup> (BRT) method, we searched for associations between known occurrences of the virus and its reservoir and the environmental conditions at those sites. As inputs for our models, we used occurrence records collated from online databases and the literature and environmental data obtained from the Inter-Sectoral Impact Model Intercomparison Project phase 2b (ISIMIP2b)<sup>59</sup>. To assess how each factor contributed to our models, we calculated their relative importance (RI). In the case of BRT models, RI is evaluated based on the number of times the factor is selected for splitting a tree, weighted by the squared improvement to the model as a result of each split, averaged over all trees<sup>59</sup>.

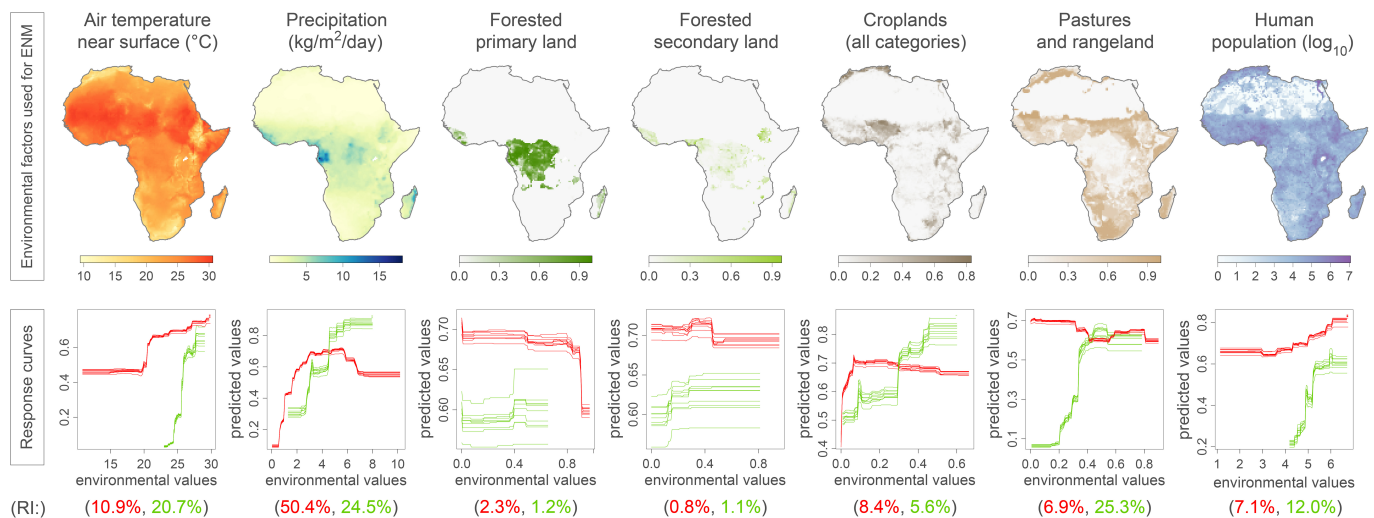
We found that for Lassa virus, three main factors contributed to the models: temperature (RI = 20.7%), precipitation (RI = 24.5%), pastures and rangeland coverage (RI = 25.3%; Fig. 1). For *M. natalensis*, we found that precipitation was the main contributor (RI = 50.4%; Fig. 1). These findings suggest that temperature, precipitation and the presence of pastures/rangeland may be the main factors influencing ecological suitability for Lassa virus, but not its reservoir species, *M. natalensis*, for which only precipitation appears to be critical.

To assess the relationship between each of our environmental factors and ecological suitability, we plotted response curves, which show how ecological suitability varies with one specific factor, while all others are kept constant at their mean. Ecological suitability values vary between 0 (unsuitable conditions) and 1 (highly suitable conditions). We found that temperatures below 25 °C or values of pastures and rangeland coverage below 20% seem unsuitable for the virus (ecological suitability < 0; Fig. 1) but still appear relatively suitable (ecological suitability > 0.4) for its reservoir species. These results indicate that even if *M. natalensis* may be found in areas with mean daily temperatures below 25 °C and limited pastures and rangeland coverage, Lassa virus is not likely to be present.

### Likely expansion of the range suitable for Lassa virus

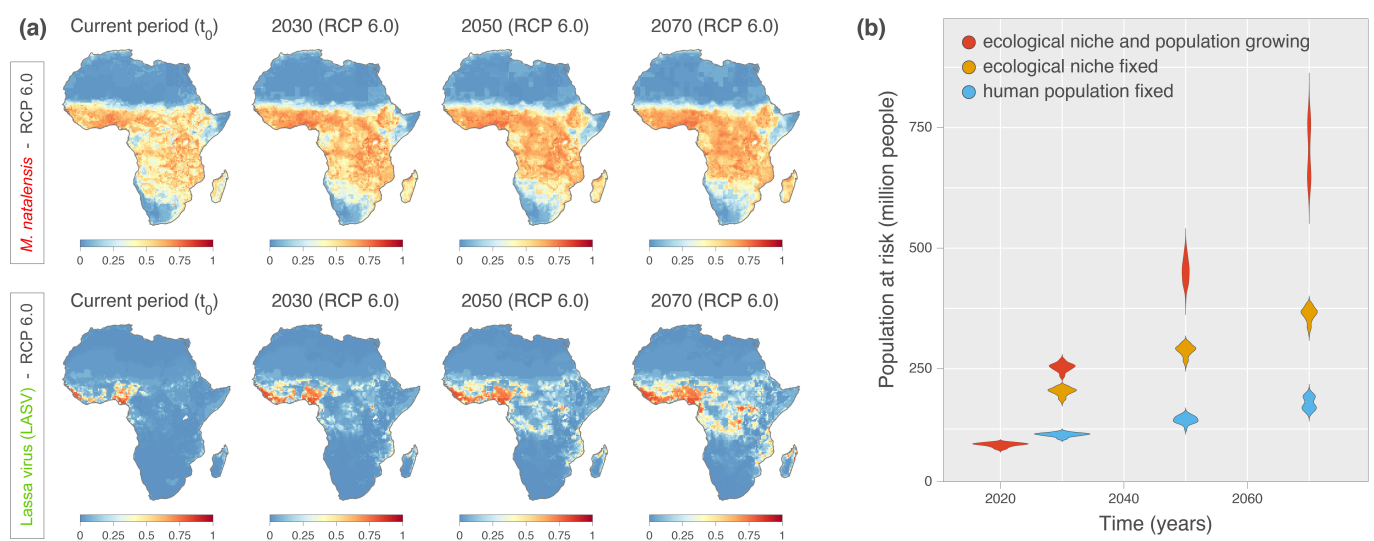
Our ecological niche modelling analyses showed that temperature, precipitation, and pastures/rangeland coverage are the main factors influencing ecological suitability for Lassa virus circulation. Due to climate change and increasing human pressure on land resources caused by population growth, these variables are expected to change in the next decades<sup>54–56</sup>. With these expected transformations, the overall area suitable for Lassa virus—also called the ecological niche of the virus<sup>60</sup>—will likely undergo substantial changes and expand. To investigate this, we used climate and land cover projections from the year 2030 to 2070 to estimate the future ecological suitability for the virus across Africa. We found that the ecological niche of Lassa virus will likely expand as new regions become suitable, notably in Central and East Africa.

We used our ecological niche models to identify the areas suitable for Lassa virus throughout Africa, based either on current or projected values of temperature, precipitation, land cover, and human population from the ISIMIP2b<sup>59</sup>. We considered environmental values projected at three-time points (2030, 2050, and 2070) according to three climate scenarios: representative concentration pathways (RCPs) 2.6, 6.0, and 8.5—which describe the evolution of global warming depending on different trajectories of greenhouse gases atmospheric concentrations<sup>61</sup>. For the present-day situation, our ecological niche maps for Lassa virus and *M. natalensis* were in agreement with previous estimates<sup>34,48</sup>, showing Lassa virus suitability across West Africa,



**Fig. 1 | Environmental factors included in the ecological niche modelling (ENM) analyses of *Mastomys natalensis* and *Lassa virus*, and their corresponding ENM response curves.** Response curves and relative importance (RI) obtained for the ENM analyses of *M. natalensis* and *Lassa virus* are coloured in red and green, respectively. The ten response curves reported for each ENM analysis correspond to ten independent repetitions of the boosted regression trees (BRT) analysis.

These response curves indicate the relationship between the environmental values and the response (i.e., the ecological suitability of *M. natalensis* or *Lassa virus*). In addition to the seven environmental factors displayed in this figure, two additional factors were also included in the ENM analyses, the non-forested primary land, and non-forested secondary land.



**Fig. 2 | Projected ecological niche suitability of *Mastomys natalensis* and *Lassa virus*, as well as human population at risk of exposure to *Lassa virus*.** **a** Projected ecological niche suitability of *Mastomys natalensis* (*M. natalensis*) and *Lassa virus* for the current period, 2030, 2050, and 2070. Each future projection (i.e., for 2030, 2050, and 2070) was performed according to four different bias-adjusted global climate models and three different representative concentration pathways (RCPs), i.e., greenhouse gas concentration scenarios considered by the Intergovernmental Panel on Climate Change (IPCC): RCP 2.6, RCP 6.0, and RCP 8.5. Here, we only report the projections obtained under RCP 6.0 (see Fig. S1 for the other scenarios as well as for the standard deviations associated with all projections, and see Fig. S2 for explicit differences between current and future

projections). For a specific time period, we report ecological niche suitability averaged over the projections obtained with the four different climatic models (see the text for further detail). **b** Projections of the human population at risk of exposure to *Lassa virus* for the current period, 2030, 2050, and 2070. For those estimations, we also re-estimate these projections while fixing the human population, i.e., not using the future projections of human population to estimate the number of people at risk (see also Fig. S3 for spatially-explicit estimation of future human exposure to *Lassa virus* and Fig. S6 for the estimations of the human population at risk of exposure to *Lassa virus* under all RCP scenarios). Source data are provided as a Source Data file.

predominantly in current endemic countries of Guinea, Sierra Leone, Liberia, and Nigeria (Fig. 2a).

At future time points, we found that the ecological niche of *Lassa virus* may substantially expand under both RCP 6.0 and RCP 8.5 (Fig. 2a, Fig. S1). RCP 2.6 and RCP 8.5 are the most extreme scenarios and refer to either stringent mitigation (RCP 2.6), or high-end emissions (RCP 8.5), while RCP 6.0 represents a medium-high emission scenario<sup>59</sup>. Focusing on RCP 6.0, we projected that by

2070, most of the region between Guinea and Nigeria will become suitable (ecological suitability >0.5) for *Lassa virus* (Fig. 2a, see Fig. S1 for the other scenarios as well as for the standard deviations associated with all projections). In addition, we found that several regions will likely become suitable in Central Africa, including in Cameroon and the Democratic Republic of the Congo (DRC), but also in East Africa, notably in Uganda. For *M. natalensis*, we found that irrespective of the scenario, the ecological niche will likely



remain stable in range, with suitability values that increase over time and across the entire niche (Fig. 2a and Fig. S1). These results show that, considering a medium-high scenario of evolution of global warming (RCP 6.0), the ecological niche of Lassa virus may expand well beyond current endemic countries, notably into parts of Central and East Africa.

To investigate the factor(s) driving the expansion of the niche of Lassa virus, we represented, on a map of Africa, the environmental values for the main factors influencing ecological suitability at current and future time points (Figs. S4, S5). In Central and East Africa, areas showing an increased suitability for the virus under RCPs 6.0 and 8.5 also exhibited an increase in temperature and pastures/rangeland land coverage (Fig. 2a, Figs. S4, S5). Based on our observations, these two factors may thus primarily drive the expansion of the range suitable for Lassa virus.

### A predicted increase in the population living in Lassa virus suitable areas

Our projections show that the potential expansion of the ecological niche of Lassa virus would occur in regions where *M. natalensis* is (and will remain) present, as indicated by our estimates of present and future suitability for *M. natalensis* across Africa (refs. 34,46–48 and Fig. 2). In the next decades, with *M. natalensis* being present where ecological conditions will be suitable for virus circulation outside of the current endemic range, the distribution of Lassa virus in Africa could potentially widen. The recent emergence of Ebola virus in West Africa and of West Nile virus in North America illustrate how zoonotic viruses can travel over long distances to effectively settle in new regions<sup>50–54</sup>, dramatically widening their range, and putting more people at risk of infection.

We next investigated how the expansion of the ecological niche of Lassa virus could affect the future number of people at risk of infection. To estimate the current and future human population in the virus niche, we considered population projections in areas with an estimated ecological suitability above 0.5 (Fig. S3). We focused again on three future time points (2030, 2050, and 2070), and three climate scenarios (RCPs 2.6, 6.0, and 8.5). We found that under RCP 6.0, the human population living in the niche of Lassa virus, where conditions are suitable for virus circulation, may increase from 92 million today (95% highest posterior density [HPD] interval: [83–98]) to 453 [414–497] million by 2050, and to 700 [624–779] million by 2070 (Figs. 2b and S6, Table S1). This increase however, may be driven by demographic growth in current suitable areas rather than by the spatial expansion of the virus ecological niche<sup>48</sup>. To investigate this, we first examined current population numbers in areas suitable for Lassa virus in 2070 (scenario RCP 6.0) and found that they are currently home to 179 million people [159–199]. This result suggests that the population is expected to grow substantially throughout the entire niche of the virus (as projected in 2070), which will more than triple by 2070 (Fig. 2b, Table S1). When comparing the number of people that will live in current or future parts of the niche in 2070, we found that population growth should be comparable in both areas (Table S1). More specifically, our results show that by 2070, 363 [333–384] million people may be exposed to Lassa virus infection in current suitable areas and that expansion of the ecological niche of the virus might put 337 [260–405] million more people at risk of infection. As our estimates may be sensitive to the arbitrary cut-off value of 0.5 we used to define an area as ecologically suitable, we repeated our analysis considering different arbitrary cut-off values (0.25 and 0.75), and observed the same trends (Table S1).

### Lassa virus circulation is remarkably slow in endemic areas

In our ecological niche modelling analyses, we found that within a few decades, ecological conditions will be suitable for Lassa virus

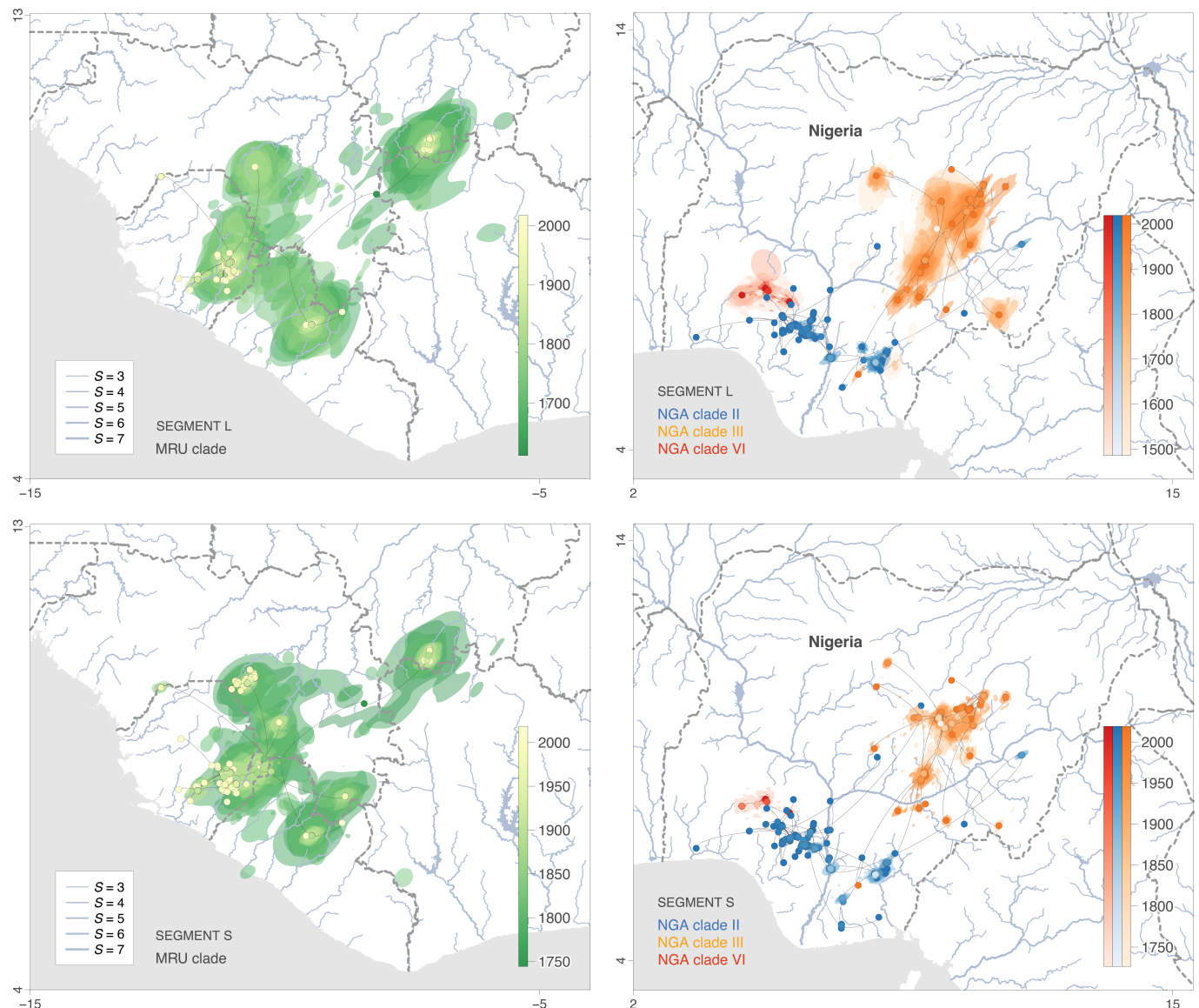
circulation beyond its current endemic range in West Africa. If Lassa virus is introduced into a new suitable region, we estimated that tens to hundreds of millions more people may be at risk of infection. The virus may however need time to spread locally and occupy a substantial area. To assess how fast the virus may spread following a potential future introduction into a suitable environment, we analysed the spatiotemporal spread of the virus using geotagged viral genomes. We showed that Lassa virus dispersal in endemic areas is remarkably slow compared to other zoonotic viruses.

To infer the spatiotemporal spread of Lassa virus since the emergence of the four major clades<sup>32</sup>, we analysed publicly available genomic sequences associated with a sampling date and location using a spatially-explicit Bayesian phylogeographic approach<sup>62</sup>. The genome of Lassa virus is segmented into a large (L) and a small (S) segment that may reassort during coinfections in *M. natalensis*<sup>63,64</sup>. As reassortment may result in distinct evolutionary histories for the L and S segments, we analysed them in separate phylogeographic inferences ( $n = 255$  and 411, respectively). We also divided our analyses between four main clades (Fig. S7): the “MRU clade” groups the subclades circulating in the Mano River Union (MRU) and Mali (also called lineages IV and V); “NGA clade II”, “III” and “VI” correspond to the main clades circulating in Nigeria (also called lineages II, III and VI, respectively)<sup>65</sup>. The trees inferred by our phylogeographic analyses capture the spatiotemporal spread of the virus (Fig. 3), with each branch representing dispersal between an estimated start and end location, and associated with an estimated duration.

To investigate the spatial spread of the main clades, we represented the trees inferred by our phylogeographic analyses on maps, separately for the MRU and the Nigerian clades. We observed that in Nigeria, the main clades are confined to distinct areas: clades II and III circulate south and north of the Niger and Benue rivers, respectively, while clade VI is limited to states in the south west (Osun, Ekiti, Ondo, Kwara, Fig. 3). We also noted that sequences from the MRU clade grouped in three main clusters circulating respectively in eastern Sierra Leone, Guinea and Mali (Fig. 3). The strong geographic structure we observed in our phylogenetic trees was consistent across the S and L segments (Figs. 3, S9, and S10) and aligns with previous studies<sup>30,31,66</sup>. These findings suggest that, although the spread of Lassa virus encompasses hundreds of years (Figs. 3, S9, and S10), virus diversity is distinct across different areas.

To approximate how fast Lassa virus circulates in endemic areas (Mano River Union and Nigeria), we estimated the weighted lineage dispersal velocity<sup>67</sup>, which corresponds to the total distance covered by the dispersal events in our trees divided by the sum of their durations. We found that Lassa virus circulates with a weighted lineage dispersal velocity between 0.8 and 1.0 km/year (95% HPD interval for the S and L segments: [0.7–1.0] and [0.9–1.0]; Table 1). Consistently across the L and S segments, our estimates of the weighted lineage dispersal velocity for each of the main clades show that virus circulation is slowest for the MRU clade and fastest for the Nigerian clade II (Fig. S8). While we cannot exclude the hypothesis that one or several clades are associated with an increased transmissibility or shorter serial intervals accelerating virus dispersal, the generally low clade-specific estimates (<1.5 km/year) remain similar. These results highlight that Lassa virus circulation is slow, which may in part explain why the main clades are confined to different areas within overall suitable regions (Fig. 2).

To determine how slow the velocity of Lassa virus circulation was compared to other zoonotic viruses, we assembled and sorted all published estimates of weighted lineage dispersal velocities (Table 1). We found that Lassa virus exhibits the slowest lineage dispersal velocity after Nova virus, while Ebola virus appeared to be the fastest. Our results indicate that Lassa virus circulation in endemic areas is particularly slow compared to other zoonotic viruses, potentially due to the small scale of the movements of its reservoir<sup>25,33</sup>.



**Fig. 3 | Spatiotemporal diffusion of Lassa virus lineages in the western Africa region and Nigeria. Maximum clade credibility (MCC) tree obtained by continuous phylogeographic inference based on 1000 posterior trees.** A separate phylogeographic analysis was performed for segments L and S as well as, in the case of the Nigerian data set, on clades II, III, and VI. These MCC trees are superimposed on 80% highest posterior density (HPD) intervals reflecting phylogeographic uncertainty. Nodes of the trees, as well as HPD regions, are coloured according to their time of occurrence, and oldest nodes (and corresponding HPD regions) are here plotted on top of youngest nodes. The trees are superimposed on

maps displaying the main rivers present in the study area and classified according to their Strahler number  $S$ , which measures the importance of a river by looking at the number of upstream rivers connected to it. International borders are represented by grey dashed lines. See also Figs. S9 and S10 for visualisations clade by clade. The “MRU clade” groups the subclades circulating in the Mano River Union (MRU) and Mali (also called lineages IV and V); “NGA clade II”, “III” and “VI” correspond to the main clades circulating in Nigeria (also called lineages II, III and VI, respectively).

### Limited impact of environmental factors on Lassa virus dispersal dynamics

Our phylogeographic inferences show that Lassa virus circulation is remarkably slow in endemic areas, which may explain, at least in part, why the spatial spread of the main clades is limited, even within overall suitable regions. This finding suggests that, in case of introduction into a new suitable area, Lassa virus may require time to spread locally. In addition, since Lassa virus depends on *M. natalensis* for transmission, any environmental feature limiting the mobility of the reservoir may also impact virus dispersal. Main waterways in particular, have been proposed to act as barriers preventing the spread of Lassa virus, based on phylogenetic evidence that virus diversity is distinct across different sides of the Niger and Benue rivers in Nigeria<sup>30,66</sup> (Fig. 3). For other viruses, such as rabies, there is evidence that environmental factors

including elevation or croplands coverage have an impact on virus lineage dispersal velocity<sup>68,69</sup>.

To determine if main waterways act as barriers to virus dispersal, we investigated whether Lassa virus tended to avoid crossing rivers based on our phylogeographic reconstructions. Using a least-cost path algorithm<sup>70</sup>, we computed the cost for the virus to travel through a landscape crossed by rivers based on both stream network data (Table S6) and the virus dispersal trajectory. We compared the cost of the observed spread inferred by our phylogeographic analyses to the cost computed under a null dispersal model that is unaware of rivers, and then estimated the statistical support for our test by approximating a Bayes Factor (BF) in favour of a cross-avoiding behaviour. We repeated our test for a range of stream sizes considering different threshold values of the Strahler number ( $S$ ) – a proxy for river stream

**Table 1 | Comparison of lineage dispersal velocities estimated for different data sets**

Data set	Weighted lineage dispersal velocity	Sampled sequences	Reference
Nova virus (moles), Belgium	0.3 km/year [0.3, 0.4]	100	Laenen et al. <sup>57</sup>
Lassa virus, segment L, Africa	0.8 km/year [0.7, 1.0]	254	(Present study)
Lassa virus, segment S, Africa	1.0 km/year [0.9, 1.0]	410	(Present study)
Rabies virus (skunks), USA	9.4 km/year [8.3, 10.6]	241	Kuzmina et al. <sup>68</sup>
Rabies virus (raccoons), USA	11.8 km/year [9.6, 13.3]	47	Biek et al. <sup>69</sup>
Rabies virus (bats), eastern Brazil	12.5 km/year [7.8, 20.3]	41	Vieira et al. <sup>70</sup>
Rabies virus (dogs), northern Africa	16.8 km/year [14.0, 19.7]	250	Talbi et al. <sup>71</sup>
Rabies virus (bats), Peru	17.7 km/year [14.6, 21.1]	260	Streicker et al. <sup>72</sup>
Rabies virus (mainly dogs), Iran	18.1 km/year [16.3, 20.8]	105	Dellicour et al. <sup>73</sup>
Rabies virus (bats), Argentina	34.7 km/year [28.1, 41.6]	131	Torres et al. <sup>74</sup>
H5N1 virus, Mekong region	149.0 km/year [115.9, 170.2]	320	Dellicour et al. <sup>75</sup>
West Nile virus, North America	165.0 km/year [158.0, 169.2]	801	Dellicour et al. <sup>76</sup>
Yellow fever virus, Brazil	169.4 km/year [131.7, 214.4]	99	Hill et al. <sup>77</sup>
Porcine deltacoronavirus, China	184.7 km/year [134.7, 234.4]	97	He et al. <sup>78</sup>
Ebola virus, West Africa	598.1 km/year [556.4, 635.3]	722	Dellicour et al. <sup>79</sup>

For each data set, we report both the posterior median estimate and the 95% highest posterior density (HPD) interval in kilometres per year (km/year).

size, based on a hierarchy of tributaries<sup>71</sup>. We found only moderate evidence (3 <BFs <20)<sup>72</sup> that the virus dispersal trajectory tends to avoid crossing rivers, no matter the Strahler number used as cut-off value for the stream size (Table S2). Overall, our results provide no strong evidence that waterways may act as notable barriers to the dispersal of Lassa virus.

We next examined how environmental conditions may affect the velocity of Lassa virus circulation considering a set of nine environmental factors for which we collected geo-referenced data from public databases (Fig. S11, Table S6). For all virus dispersal events inferred by our phylogeographic analyses, we investigated whether the duration of the dispersal correlated with the environmental factors in our testing set. To assess these correlations, we computed an “environmental distance”, which corresponds to the distance of the dispersal event, weighted according to the environmental conditions along the path of dispersal. Our procedure only considers constant-in-time environmental values that do not reflect the climatic and land cover conditions during the earliest part of Lassa virus dispersal history, so we restricted our analyses to the most recent dispersal events (corresponding to tip branches of the trees from our phylogeographic reconstructions; Fig. S7). On average, this corresponds to considering lineage dispersal events starting in 1991 for segment L (95% HPD: [1988–1995]) and 1993 for segment S (95% HPD: [1991–1994]). We only found moderate evidence (3 <BFs <20) that the presence of savannas may slow down viral circulation (Tables S3 and S4). These results suggest that the environmental factors considered in our analysis have no dramatic impact on the velocity of Lassa virus circulation.

### Limited virus propagation following introduction into a new suitable area

In the post hoc analyses of our phylogeographic inferences, we did not identify any environmental factor that may prevent or notably slow down virus spread in a suitable environment. Hence, in case of introduction into a new suitable region, the main parameter that we can expect to limit Lassa virus propagation based on our analysis would be its slow lineage dispersal velocity. To illustrate how a slow lineage dispersal velocity may limit the spatial extent of virus spread following a potential introduction, we simulated virus dispersal based on the parameters inferred by our phylogeographic analyses (Fig. 4). We ran simulations over a 20-year period in two areas: one projected to become suitable for virus circulation by 2050 under scenario RCP 6.0 and the other one, under RCP 8.5. To simulate virus dispersal, we randomly sampled dispersal events inferred by our phylogeographic

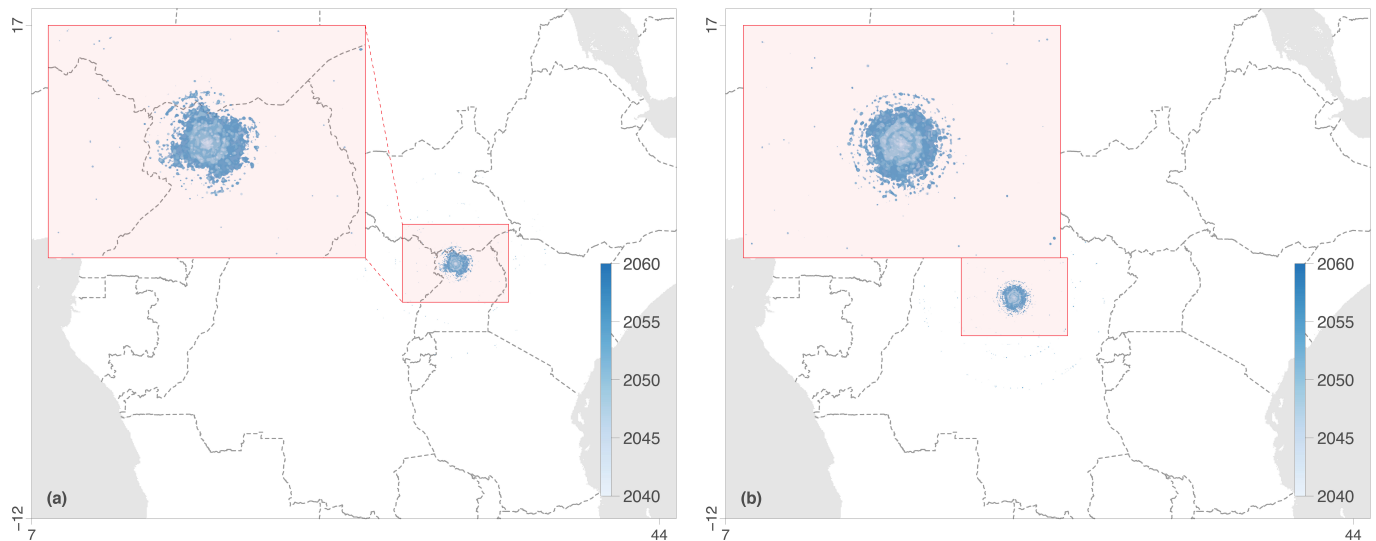
analyses for the Nigerian clade II, for which we have the largest number of sequences. We set the trajectory of dispersal events by selecting the ending location with a probability equal to the local ecological suitability (as projected in our ecological niche modelling analyses). By mapping the results of 1000 simulations of virus dispersal, we show that Lassa virus would likely remain confined within a range of ~200 km<sup>2</sup> (Fig. 4), even when starting within a large suitable area (e.g., with scenario RCP 8.5, Fig. 4 and Fig. S14). Our simulations show how, if Lassa virus circulation is as slow as in current endemic areas, virus propagation would remain spatially limited over the first decades following its introduction into a new ecologically suitable area.

## Discussion

Previous molecular dating studies have shown that Lassa virus has been circulating for at least 1000 years and originated in present-day Nigeria, from where it spread to the West, reaching into the MRU region<sup>14,32,73,74</sup>. Lassa virus is considered endemic in Guinea, Liberia, Nigeria, and Sierra Leone, but the virus likely circulates in other neighbouring countries along its presumed dispersal path<sup>14–17</sup>. Although we do not attempt to precisely map the current range of Lassa virus, our ecological suitability estimates for the current period appear globally similar to the results obtained in an earlier study<sup>34</sup> and show areas suitable for virus circulation across West Africa (see Fig. 2a). Consistent with reports of Lassa virus infections in humans and rodents outside of endemic hotspots, our results suggest that the virus may be present in most coastal West African countries and Mali, prompting strengthened Lassa fever surveillance throughout the whole region.

*M. natalensis* is considered the primary reservoir of Lassa virus<sup>24</sup> and it is still unclear why the distribution of this rodent species extends far beyond that of the virus, which is limited to West Africa<sup>33</sup>. Our analyses show that different environmental factors determine ecological suitability for the virus and its host, suggesting that the absence of Lassa virus beyond West Africa could be partly due to environmental constraints. To be able to estimate future ecological suitability for Lassa virus circulation, we have used projected environmental data with a low resolution (0.5 decimal degrees), due to the coarse scale of climate change projections<sup>59,75</sup>. This limited resolution reduced our ability to account for small-scale environmental variations that could affect suitability for Lassa virus; however, the good performance of our models (area under the receiver operating characteristic curves between 0.74 and 0.85) suggests that our approach provides a reasonable estimate of the distribution of Lassa virus infections.





**Fig. 4 | Phylogeographic simulations of viral spread following a successful introduction into a new ecologically suitable area.** Phylogeographic simulations are based on the tree topologies inferred for Lassa virus clade II (segment S). Virus dispersal was constrained by ecological suitability using ecological niche projections for Lassa virus in 2050 according to scenarios RCP 6.0 (a) and RCP 8.5 (b).

95% highest posterior density (HPD) polygons are coloured according to time and based on 1000 simulations starting from the same ancestral location. For each set of simulations, a zoom on the outcome is shown. For the illustration, five distinct phylogeographic simulations per scenario are also displayed in Fig. S14.

In addition to environmental aspects, several other factors may also contribute to the difference in distribution between Lassa virus and its reservoir. As reported for other mammarenaviruses, the virus may only be present in the *M. natalensis* subtaxon A-I<sup>36,37</sup>. Records of Lassa virus infection in *M. natalensis* subtaxon A-II<sup>76</sup> and in other rodent species including *M. erythroleucus* or *Hylomyscus pamfi*<sup>77</sup> suggest, however, that susceptibility to Lassa virus infection may not be species or subtaxon specific. Other possible explanatory factors include intra-host competition between different viruses<sup>38,78</sup>, or cross-immunity due to the circulation of closely related viruses<sup>40,41</sup>. As most other old world arenaviruses that circulate in *M. natalensis* are found only in East Africa<sup>12,36,37</sup>, there is little data to assess these two mechanisms based on field data. Only in Mayo-Ranewo (Taraba state, eastern Nigeria), rodent trapping studies have identified a Mobala-like virus in *M. natalensis*<sup>76,79</sup>, which does not seem to effectively restrict the transmission of Lassa virus, as human infections are often reported in that state of Nigeria<sup>80</sup>.

In our phylogeographic analyses we find that Lassa virus mostly spreads on a small spatial scale, with relatively few long-distance dispersal events (Fig. 3) but we do not identify environmental factors that seem to strongly restrict or slow down virus spread. Using a phylogeographic simulation procedure, we also show that a slow lineage dispersal velocity would likely result in a limited spatial propagation if Lassa virus was successfully introduced in a new ecologically suitable area. The slow spread of Lassa virus may be due to the small scale of the movements of its reservoir, as suggested by genetic studies showing that *M. natalensis* rodents travel rarely outside of their commensal habitat and are prone to high levels of consanguinity<sup>25,33</sup>. However, it is surprising that Lassa virus — and by implication its reservoir — seem unrestricted by the environment. Of note, our results were not always consistent between the L and S segments, possibly due to the lower number of L sequences (255) in our data set compared to S sequences (411). More generally, the number of genomic sequences in our datasets may offer limited power for the tests we used to assess the possible impact of rivers and other environmental factors on virus spread. Hence, a larger sampling of Lassa virus genomes throughout the virus range would allow for better evaluation of the role of the environment in limiting spread. Another caveat here is that our procedure considers

constant-in-time environmental values that do not reflect changes in climatic and land cover conditions that may have occurred throughout the virus dispersal history. Indeed, it is estimated that in the past century alone, about 10 million ha of forest were lost in West Africa, mostly to the benefit of agricultural expansion<sup>81</sup>. Over the same period, annual precipitation has decreased throughout the area while temperatures have risen, up to over 1 °C in some parts of Nigeria<sup>82</sup>. To address this issue, we restricted our analysis to the most recent dispersal events in the virus dispersal history which date back, on average, to the 1990s. However, our approach still does not account for abrupt changes in climate, land use, and population density that may have occurred throughout the last decades in the West African region<sup>56</sup>.

In our study, we use phylogeographic simulations to highlight how, in the absence of restrictions from the environment, a slow lineage dispersal velocity may limit the propagation of Lassa virus in case of a successful introduction into a new ecologically suitable area. We use these simulations for illustration and not prediction as the dispersal dynamics upon virus emergence in a new region are unclear. The virus may spread swiftly through an immunologically naïve rodent population, but the low mobility of the rodent reservoir could still limit the velocity of virus dispersal on a larger scale. As underlined above, a number of other elements may also come into play, such as the nature of host species or subtaxa as well as potential cross-immunity or competition due to the local co-circulation of closely related viruses. Nevertheless, it is worth pointing out that following the emergence of Lassa virus in the MRU, virus circulation remained as slow — if not slower — as in Nigeria, as highlighted by our estimates of weighted lineage dispersal velocity (Fig. S8).

Our ecological niche modelling analyses highlight a risk of expansion of Lassa virus towards regions in Central and East Africa that could potentially lead to a drastic increase in the number of people living in areas ecologically suitable for Lassa virus (Table S1). To reach the largest ecologically suitable regions we identify in DRC and Uganda, the virus would have to spread over several hundreds of kilometres and cross regions with low ecological suitability. Such long-distance movements likely allowed Lassa virus to reach the Mano River Union from Nigeria several hundred years ago<sup>32</sup>. This early part of the virus dispersal history, however, remains poorly understood and it is

thus hard to predict if the virus is likely to travel across the African continent again. In addition, some factors may hinder virus dispersal following a hypothetical introduction into a new ecologically suitable region. First, in our analyses, as there is evidence that Lassa virus infection may not be species or subtaxon-specific<sup>76,77</sup>, we considered the entire *M. natalensis* species to be susceptible to Lassa virus infection. It is, however, possible that the virus spreads less (or more) efficiently in a different subtaxon of *M. natalensis*. Second, while in most of the areas where ecological suitability for LASV appears to increase over time outside of West Africa (e.g., DRC, Uganda, Cameroon), no *Mammarenavirus* species have so far been identified<sup>12</sup>, we cannot rule out the possibility that potential cross-immunity or competition due to the local co-circulation of closely related viruses may hinder the propagation of LASV.

To provide a very conservative estimate of the future risk of exposure to Lassa virus, we can focus on population growth in the endemic range and leave aside its possible expansion. When considering a cut-off value of 0.5 for the ecological suitability threshold defining the niche of the virus, we estimated that population growth in endemic countries could alone put 186 (95% HPD interval: [172–196]) and 341 [315–369] million people at risk of infection by 2030 and 2070, respectively (Table S1), compared to an estimated 83 [76–87] million today. A limitation to these estimates is that our population projections do not take into account migrations due to environmental and climate change pressures, which could affect projections in regions where extreme weather conditions are expected.

In this analysis, we focused on evaluating the population that will likely be exposed to Lassa in endemic regions or that might be exposed in regions that we predict to be suitable for Lassa virus in the future. Importantly, we did not evaluate how much heightened surveillance, improved sanitation, and increased awareness may help to reduce exposure to Lassa virus in the future<sup>83–86</sup>. These measures will, however, remain crucial to prevent infections. Even more so as we report an extremely slow dispersal velocity of Lassa virus in endemic areas, suggesting that localised efforts that target infection hotspots may prove to be highly effective.

A large part of the population growth expected in endemic areas is driven by Nigeria (-91%), a country that has reported an unusual increase in the number of reported Lassa fever cases over the last two years<sup>80,87</sup>. This uptick was not attributed to increased inter-human transmission<sup>30,31</sup> or to the emergence of a specific viral strain<sup>30,31</sup>; but raised the question of a more intense circulation within the reservoir or of an improvement in surveillance and public awareness. To discriminate between these two hypotheses, we investigated the evolution of the overall genetic diversity of Lassa virus in the main Nigerian clade (clade II) over the past decades, using a coalescent approach that accounts for preferential sampling. We found that the effective population size of Nigeria clade II increased over the last years (segment S; Fig. S13), suggesting that the recent uptick in cases in Nigeria was not the sheer result of an improvement in surveillance. Hence, even if Lassa virus does not expand to new regions in the near future, the virus still actively circulates in increasingly populated endemic areas, and there is thus an urgent need for more efficient prophylactic and therapeutic countermeasures.

With anthropogenic climate change and an increasing impact of human activities on the environment, extensive studies of the ecology and spread of zoonotic and vector-borne diseases are needed to anticipate possible future changes in their distribution<sup>88,89</sup>. We showed that changes in temperature, precipitation, and pastures/rangeland land coverage may expand the ecological niche of Lassa virus beyond current endemic areas, potentially exposing hundreds of million more people to Lassa. By simulating virus spread, we highlight that, in the scenario of a successful introduction and propagation of the virus in a new ecologically suitable area, the emerging circulation foci could remain

limited to a small spatial scale over the first decades. Our study provides an example of how ecological niche modelling and spatially-explicit phylogeography can be effectively combined to investigate the future risk of a major zoonotic disease.

## Methods

### Ecological niche modelling of *Mastomys natalensis* and Lassa virus

We employed the boosted regression trees<sup>90</sup> (BRT) approach implemented in the R package *dismo*<sup>88</sup> (version 1.3-8) to perform ecological niche modelling analyses of both Lassa virus and its host, *M. natalensis*. BRT is a machine learning method that allows to model complex non-linear relationships between the probability of occurrence and various predictor variables<sup>88,91</sup>. This approach aims to generate a collection of sequentially fitted regression trees that optimise the predictive probability of occurrence based on predictor values<sup>90,91</sup>, which can also be interpreted as a measure of ecological suitability. In a comprehensive review of distribution modelling methods, Elith et al.<sup>90</sup> found BRT to perform best along with the maximum entropy method<sup>92</sup>.

The BRT approach requires both presence and absence data. When unavailable, as this is the case for Lassa virus and its host, absence data can be approximated by random pseudo-absence points sampled from the study area (also referenced as the “background”). For Lassa virus, we only sampled pseudo-absences in raster cells in which the presence of *M. natalensis* has been recorded. While heterogeneous disease surveillance efforts likely bias the spatial distribution of Lassa records<sup>46,47</sup>, this procedure avoids treating under-sampled areas as ecologically unsuitable for the virus, but also limits the potential impact of such heterogeneity in sampling effort or surveillance<sup>93–95</sup>. Of note, while several rodent species such as *M. erythroleucus*<sup>77</sup> may be important for Lassa virus transmission, we chose to be conservative and focused on the main species recognised as a reservoir host for Lassa virus to define our background. Similarly, for *M. natalensis*, we only sampled pseudo-absences in raster cells in which the presence of at least one individual of another species belonging to the Muridae family has been recorded. Because it only requires a single occurrence record to consider a presence, we discarded all but one occurrence record per raster cell. We applied the same filtering step for the pseudo-absence points and simply discarded pseudo-absences falling in raster cells with occurrence data. This filtering procedure is required to only have a presence or a pseudo-absence assigned to each grid cell, but also has the advantage of minimising the risk of artefacts due to spatially heterogeneous disease/wildlife surveillance efforts in the resulting BRT models. Indeed, even in areas where the disease surveillance was limited, it only necessitates at least a single occurrence record to consider the presence of the target species in the large grid cell we use.

To select the optimal number of trees in the BRT models, we used a spatial cross-validation procedure based on five spatially separated folds generated with the *blockCV* R package<sup>95</sup> (version 2.1.4). We employed a spatial rather than a standard cross-validation because the latter may overestimate the ability of the model to make reliable predictions when occurrence data are spatially auto-correlated<sup>97</sup>, which can frequently be the case. All BRT analyses were run and averaged over 10 cross-validated replicates, with a tree complexity set at 5, an initial number of trees set at 100, a learning rate of 0.005, and a step size of 10. We evaluated the inferences using the area under the receiver operating characteristic (ROC) curve, also simply referred to as “area under the curve” (AUC). Among replicates, AUC values ranged from 0.68 to 0.73 for *M. natalensis* (mean = 0.71), and from 0.74 to 0.85 for Lassa virus (mean = 0.79).

We obtained occurrence data for *M. natalensis* species from the Global Biodiversity Information Facility (GBIF.org, <https://doi.org/10.15468/dl.hrjyjl>, accessed 2019-07-19), the Integrated Digitized Biocollections (<https://www.idigbio.org>, accessed 2020-01-04), the



Field Museum of Natural History Zoological collections (<https://collections-zoology.feldmuseum.org>, accessed 13 December 2019), and the African Mammalia database (<http://projects.biodiversity.be/africanmammalia>, accessed 2019-12-14). This data set was supplemented with the data available in the scientific literature (search for term “*Mastomys natalensis*”, in PubMed and Google). Duplicate records as well as records located in the ocean were excluded from the final data set, totalling 2504 unique *M. natalensis* occurrence records. For 26 of those records, the location was not provided as spatial coordinates but as a locality (below or at the administrative level 4). Therefore, the latitude and longitude data correspond to that of the locality (determined as described in the subsection *Selection and preparation of viral sequences*; see below). Occurrence data for the *Muridae* family were obtained from the GBIF database (GBIF.org, <https://doi.org/10.15468/dl.cs3c41>, accessed 2019-07-19). Duplicate records and records located in the ocean were excluded from the data set, totalling 10,806 unique *Muridae* occurrence records for the African continent. Occurrence data for Lassa virus were obtained by combining the data set from Fichet-Calvet & Rogers<sup>48</sup> with records associated with sequences from our Lassa virus sequence data set (see below the subsection *Selection and preparation of viral sequences* for further detail), records of infected *M. natalensis* from our host occurrence data set and the data available in the scientific literature (search for term “Lassa virus” in PubMed). Duplicate records were discarded from the data set, resulting in 310 unique Lassa virus occurrence records. For two of those records, the location was not provided as spatial coordinates but as a locality (below or at the administrative level 4) so the latitude and longitude data corresponded to that of the locality (determined as described in the subsection *Selection and preparation of viral sequences*; see below). Our BRT models were trained on current environmental factors and then used to obtain estimates of future ecological niches for both Lassa virus and *M. natalensis*.

The BRT analyses were based on several environmental factors: harmonised present-day and future climate, land cover, and population data available through the Inter-Sectoral Impact Model Intercomparison Project phase 2b (ISIMIP2b)<sup>59</sup>. The climate information consists of daily gridded near-surface air temperature and surface precipitation fields derived from four bias-adjusted<sup>98</sup> global climate models (GCMs; GFDL-ESM2M<sup>99</sup>, HadGEM2-ES<sup>100</sup>, IPSL-CM5A-LR<sup>101</sup>, and MIROC5<sup>102</sup>) participating in the fifth phase of the Coupled Model Intercomparison Project (CMIP5<sup>103</sup>). We considered simulations conducted under historical climate forcings and RCPs 2.6, 6.0, and 8.5. In addition, we considered observed gridded temperature and precipitation from the concatenated products GSWP3 and EWEMBI<sup>58</sup> for assessing the current (1986-2005) conditions. For land cover, we used version 2 of the Land Use Harmonisation (LUH2<sup>104</sup>) providing historical and projected land cover states under a range of shared socio-economic pathways (SSPs), and from which we considered SSP1-26, SSP4-6.0, and SSP8-85. Finally, we retrieved gridded population projections<sup>105</sup> under SSP2-26. For each combination of product (GCM, GSWP3-EWEMBI LUH2, gridded population), scenario (historical, RCP, SSP), and analysis window (1986-2005, 2021-2040, 2041-2060, and 2061-2080), we computed the grid-scale temporal mean. For each scenario and time period, we estimated an index of human exposure (IHE) which corresponds to human population estimates ( $\log_{10}$ -transformed) in raster cells associated with an ecological suitability for Lassa virus above or equal to 0.5. Specifically, we used these IHE values to calculate the number of people at risk of exposure to Lassa virus. To investigate the specific effect of human population growth in current and future suitable areas, we also re-estimated future IHE values using (i) current population estimates with future projections of ecological suitability for Lassa virus to estimate population growth throughout current and future suitable areas, and (ii) future projections of human population with current projection of ecological suitability for Lassa

virus to estimate the future population living in current suitable areas (Table S1). For each estimate, we calculated the mean and 95% HPD interval across all climatic models and ecological niche model replicates.

### Selection and curation of viral sequences

All publicly available sequences for Lassa virus were downloaded from the NCBI Nucleotide database (keywords: “*lassa* NOT *mopeia* NOT *natalensis*”;  $n = 729$  L and 1202 S sequences; database accessed on October 31, 2019). They were combined with recently generated sequences from Nigeria that have been sequenced using the MinION technology (Oxford Nanopore) in conjunction with a non-targeted metagenomic RNA sequencing approach<sup>31</sup> and that are publicly available on the website virological.org (<https://virological.org/t/2019-lassa-virus-sequencing-in-nigeria-final-field-report-75-samples/291>).

We filtered the data by: (i) excluding laboratory strains (adapted, passaged multiple times, recombinant, obtained from antiviral or vaccine experiments), (ii) excluding sequences without a timestamp, (iii) keeping only sequences from a single timepoint (if multiple timepoints were available for a patient), (iv) removing duplicates (when more than one sequence was available for a single strain), and (v) excluding sequences from identified hospital epidemics or sequences for which the location corresponded to the site of hospitalisation. The remaining sequences were trimmed to their coding regions and arranged in sense orientation separately for the S segment (NP-NNN-GPC) and the L segment (L-NNN-Z). The sequences were aligned using MAFFT<sup>106</sup> (version 7) and inspected manually using the program AliView<sup>107</sup> (version 1.26). At this step, we discarded low-quality sequences (manual curation) and very short sequences (combined ORF length <500nt). Since there is an overlap between the sequence data from the work of Siddle and colleagues<sup>30</sup> and of Kafetzopoulou and colleagues<sup>31</sup>, we excluded sequences with zero or one mismatch between the two sets of sequences to ensure that there would not be duplicates in our data sets. Two types of alignments were generated. The alignments with all curated sequences regardless of the availability of detailed location information included 756 S segment sequences and 551 L segment sequences, respectively. The alignments with detailed location information included 411 S segment sequences and 255 L segment sequences, respectively. For the sequences with detailed location information, when no spatial coordinates were provided but only a name, spatial coordinates were determined using a combination of online platforms (Table S5). When several coordinates were available for one location, those matching across several data sets were kept, if the location was found in only one data set, the coordinates corresponding to the highest administrative level were kept.

### Inferring the dispersal history of Lassa virus lineages

We performed spatially-explicit phylogeographic reconstructions using the relaxed random walk (RRW) diffusion model<sup>62</sup> implemented in BEAST 1.10<sup>96</sup>, which was coupled with the BEAGLE 3 library<sup>108</sup> to improve computational performance. We modelled the nucleotide substitution process according to a GTR+ $\Gamma$  parameterisation<sup>109</sup> and branch-specific evolutionary rates according to a relaxed molecular clock with an underlying log-normal distribution. These phylogeographic analyses were based on the alignments of sequences associated with known spatial coordinates. For both the demographic and phylogeographic reconstructions, we ran a distinct BEAST analysis for each segment (L and S), sampling Markov chain Monte-Carlo (MCMC) chains every  $10^5$  generations. We used Tracer 1.7<sup>110</sup> for identifying the number of sampled trees to discard as burn-in, but also for inspecting the convergence and mixing, ensuring that estimated sampling size (ESS) values associated with estimated parameters were all >200. We used TreeAnnotator 1.10<sup>96</sup> to obtain a maximum clade credibility (MCC) tree for each BEAST analysis. Finally, we

used the R package *seraphim*<sup>67</sup> to extract the spatiotemporal information embedded within trees obtained by spatially-explicit phylogeographic inference, as well as to estimate the weighted lineage dispersal velocity.

### Impact of environmental factors on the dispersal dynamics of Lassa virus lineages

Based on the spatially-explicit phylogeographic reconstructions, we performed two different kinds of analyses to investigate the impact of several environmental factors on the dispersal history and dynamics of Lassa virus lineages. First, we tested the impact of main rivers acting as potential barriers to Lassa virus dispersal (see Table S6 for the source of the original rivers shapefile). For this purpose, we used the least-cost path algorithm<sup>70</sup> to compute the total cost for viral lineages to travel through a landscape crossed by rivers. This algorithm uses an underlying environmental raster to compute the minimum cost to move from one position to another. Here, we generated rasters by assigning a value of 1 to raster cells that were not crossed by a main river and a value of  $1 + k$  to raster cells crossed by a main river (raster resolution:  $-0.5$  arcmin). Because the raster cells that were not crossed by a main river were assigned a uniform value of 1,  $k$  thus defines the additional resistance to movement when the cell does contain such a potential landscape barrier<sup>111,112</sup>. In order to assess the impact of that rescaling parameter, we tested three different values for  $k$ : 10, 100, and 1000. Furthermore, as the notion of main river is arbitrary, we used different threshold values of the Strahler number  $S$  associated with each river to select the main rivers to consider in each analysis. In hydrology,  $S$  can be used as a proxy for stream size by measuring the branching complexity, i.e., the position of a river within the hierarchical river network. In practice, we compared the total cost computed for posterior trees with the total cost computed on the same trees along which we simulated a stochastic diffusion process under a null dispersal model<sup>113</sup>. Hereafter referred to as “simulated trees”, these trees were obtained by simulating a relaxed random walk process along the branches of trees sampled from the posterior distribution obtained by spatially-explicit phylogeographic inference<sup>113</sup>. Because this stochastic diffusion process did not take the position of rivers into account, we can expect the total cost to be lower for inferred trees under the assumption that viral lineages did tend to avoid crossing rivers. For each inferred or simulated tree, we computed the total cost TC, i.e., the sum of the least-cost values computed for each phylogenetic branch considered separately. Each inferred TC value ( $TC_{\text{inferred}}$ ) was then compared to its corresponding simulated value ( $TC_{\text{simulated}}$ ) by approximating a Bayes factor (BF) support as follows:  $BF = [p_e / (1 - p_e)] / [0.5 / (1 - 0.5)]$ , where  $p_e$  is the posterior probability that  $TC_{\text{simulated}} > TC_{\text{inferred}}$ , i.e., the frequency at which  $TC_{\text{simulated}} > TC_{\text{inferred}}$  in the samples from the posterior distribution. The prior odds is 1 because we can assume an equal prior expectation for  $TC_{\text{inferred}}$  and  $TC_{\text{simulated}}$ .

Next, we tested the impact of several environmental variables, again described as rasters, on the dispersal velocity of Lassa virus lineages (Table S6, Fig. S11): main rivers (as defined by selecting rivers with a  $S$  value higher than 2, 3, 4, 5, and 6), forest areas, grasslands, savannas, croplands, annual mean temperature, annual precipitation, and human population density. Except for the generated river rasters (see above), all these rasters had a resolution of  $-2.5$  arcmin. For one-dimensional landscape features such as rivers, we had to resort to higher resolution rasters ( $-0.5$  arcmin) to obtain sufficiently precise pixelations (rasterizations) for this non-continuous environmental factor. Raster cells assigned to rivers would otherwise be exceptionally large given the size of the study area, which could potentially lead to artefactual results. Environmental rasters were tested as potential conductance factors (i.e., facilitating movement) as well as potential resistance factors (i.e., impeding movement). For each environmental variable, we also generated several distinct rasters with the following formula:  $v_i = 1 + k^*(v_o/v_{max})$ , where  $v_i$  is the transformed cell value,  $v_o$  the

original cell value, and  $v_{max}$  the maximum cell value recorded in the raster. The rescaling parameter  $k$  here allows the definition and testing of different strengths of raster cell conductance or resistance, relative to the conductance/resistance of a cell with a minimum value set to 1<sup>114</sup>. For each of the three environmental factors, we again tested three different values for  $k$  (i.e., 10, 100, and 1000). The following procedure can be summarised in three successive steps<sup>115</sup>: (i) based on environmental rasters, we computed environmental distance for each branch in inferred and simulated trees. These distances were computed using two different algorithms: the least-cost path and Circuitscape algorithm, the latter using circuit theory to accommodate uncertainty in the route taken<sup>116</sup>. For computational tractability, high resolution river rasters were only tested with the least-cost path algorithm. (ii) We estimated the correlation between time durations and environmental distances associated with each phylogenetic branch. Specifically, we estimated the statistic  $Q$  defined as the difference between the coefficient of determination obtained when branch durations are regressed against environmental distances computed on the environmental raster, and the coefficient of determination obtained when branch durations are regressed against environmental distances computed on a uniform “null” raster, i.e., a uniform raster with a value of “1” assigned to all its cells. We estimated  $Q$  for each tree and we thus obtained two distributions of  $Q$  values: one for inferred and one for simulated trees. We only considered an environmental raster as potentially explanatory if both its distribution of regression coefficients and its associated distribution of  $Q$  values were positive<sup>117</sup>. (iii) We evaluated the statistical support associated with a positive  $Q$  distribution (i.e., with at least 90% of positive values) by comparing it with its corresponding null distribution of  $Q$  values based on simulated trees. We formalised this comparison by approximating a BF support as defined above, but this time defining  $p_e$  as the posterior probability that  $Q_{\text{estimated}} > Q_{\text{simulated}}$ , i.e., the frequency at which  $Q_{\text{estimated}} > Q_{\text{simulated}}$  in the samples from the posterior distribution<sup>68</sup>. For computational reasons, the main rivers rasters, which had to be associated with higher resolution (see above), were only tested as resistance factors with the least-cost-path algorithm.

### Phylogeographic simulations

We implemented a phylogeographic approach to simulate virus dispersal over a 20-year period following a successful introduction event within a new ecologically suitable area in 2050 under scenarios RCP 6.0 and RCP 8.5. We simulated viral lineage dispersal events by randomly sampling from the dispersal events inferred by our phylogeographic analyses. These simulations were performed under the assumption of no notable impact of underlying environmental factors. To set the trajectory of lineage dispersal events, we selected the ending location with a probability defined by the local ecological suitability. The starting point of those simulations was selected arbitrarily within the most suitable area of the extended part of the ecological niche estimated for Lassa virus in 2050, and was thus different for simulations performed under scenarios RCP 6.0 and RCP 8.5.

### Inferring the demographic history of Lassa virus lineages

We performed demographic reconstructions using the flexible skygrid coalescent model<sup>118</sup> implemented in BEAST 1.10<sup>96</sup>. The skygrid model allows to estimate the past evolution of the viral population effective size through time. For these analyses, we also modelled the nucleotide substitution process according to a GTR +  $\Gamma$  parameterisation<sup>109</sup> and branch-specific evolutionary rates according to a relaxed molecular clock with an underlying log-normal distribution<sup>110</sup>. In the case of NGA clade II for which we inferred a recent increase in the global effective population size, we also performed a preferential sampling analysis<sup>119</sup>. By modelling the sampling times as a process dependent on effective population size, this complementary analysis allows to explicitly take into account heterogeneous sampling density through time, which can improve estimates of global effective population size<sup>119</sup>.

## Reporting summary

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

## Data availability

All source data used in this study are available at [https://github.com/sdellcour/lassa\\_spreads](https://github.com/sdellcour/lassa_spreads) (<https://doi.org/10.5281/zenodo.6998624>). The sources of the different raster files used in this study are provided in Table S6. We obtained occurrence data for the Muridae family from the Global Biodiversity Information Facility (<http://www.gbif.org>, accessed 2019-07-19, GBIF occurrence downloads <https://doi.org/10.15468/dl.cs3c41>). For the *M. natalensis* species, we obtained occurrence data from the Global Biodiversity Information Facility (<http://www.gbif.org>, accessed 2019-07-19, GBIF occurrence downloads <https://doi.org/10.15468/dl.hrjy1>), the Integrated Digitized Biocollections (<http://www.idigbio.org/portal> (2020), Query: {"filtered": {"filter": {"and": [{"exists": {"field": "geopoint"}}, {"term": {"scientificname": "mastomys natalensis"}}]}}, 4348 records, accessed on 2020-01-04T05:40:40.066945, contributed by 19 Recordsets), the Field Museum of Natural History Zoological collections (Field Museum of Natural History (Zoology) Mammal Collection [https://collections-zoology.fieldmuseum.org/list?f%5B0%5D=ss\\_CatCatalog%3A%22Mammals%22&\\_ga=2.123662347.1070684726.1508778418-143671043.1493067972](https://collections-zoology.fieldmuseum.org/list?f%5B0%5D=ss_CatCatalog%3A%22Mammals%22&_ga=2.123662347.1070684726.1508778418-143671043.1493067972), accessed 2019-12-13), and the African Mammalia database (African Mammalia, <http://projects.biodiversity.be/africanmammalia/search>, accessed 2019-12-14). Full citations are provided in "Citations\_rodent\_occurrence\_data.txt" available on the GitHub repository referenced above. This data set was supplemented with the data available in the scientific literature (search for term "Mastomys natalensis", in PubMed and Google). For each record used in this analysis, the specific record or collection ID is specified in the file "Mastomys\_natalensis\_RK050820.csv" or "Muridae\_family\_allData\_RK220819.csv", both available on the GitHub repository referenced above. The sources used to retrieve sampling coordinates for Lassa virus samples are listed in Table S5. For Lassa virus, occurrence data were obtained from the scientific literature (search for term "Lassa virus", in PubMed and Google) and the source of each record used in this analysis is specified in the file "Lassa\_virus\_cases\_RK070820.csv" available on the GitHub repository referenced above. The sources of the different raster files used in this study are provided in Table S6. Data for the environmental factors used in the BRT analyses was obtained from the Inter-Sectoral Impact Model Intercomparison Project phase 2b (ISIMIP2b, <https://data.isimip.org/>). LASV sequences analysed in the present study were available on GenBank before November 20, 2019, except for the LASV sequences from cases sampled during the year 2019 in Nigeria, which are publicly available on the website virological.org (<https://virological.org/t/2019-lassa-virus-sequencing-in-nigeria-final-field-report-75-samples/291>). Accession numbers of selected genomic sequences are listed in the file "LASV\_all\_the\_metadata.csv" available on the GitHub repository referenced above. All processed data (BRT models, BRT predictions, phylogeographic inferences, dispersal statistics estimations, and seraphim analyses) generated in this study are also available on the GitHub repository referenced above. Source data are provided with this paper.

## Code availability

All R scripts and related files needed to run all the ecological niche modelling and landscape phylogeographic analyses, as well as BEAST XML files, are publicly available at [https://github.com/sdellcour/lassa\\_spreads](https://github.com/sdellcour/lassa_spreads) (<https://doi.org/10.5281/zenodo.6998624>).

## References

1. Morens, D. M. et al. The origin of COVID-19 and why it matters. *Am. J. Trop. Med. Hyg.* **103**, 955–959 (2020).
2. Pierson, T. C. & Diamond, M. S. The emergence of Zika virus and its new clinical syndromes. *Nature* **560**, 573–581 (2018).
3. Gates, B. The next epidemic—Lessons from Ebola., <https://doi.org/10.1056/NEJMp1502918> (2015).
4. World Health Organization. *Lassa fever research and development (R&D) roadmap*. [https://www.who.int/publications/m/item/lassa-fever-research-and-development-\(r-d\)-roadmap](https://www.who.int/publications/m/item/lassa-fever-research-and-development-(r-d)-roadmap) (2018).
5. World Health Organization. Prioritizing diseases for research and development in emergency contexts. <https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts>.
6. Akpede, G. O. et al. Caseload and case fatality of Lassa fever in Nigeria, 2001–2018: A specialist center's experience and its implications. *Front. Public Health* **7**, <https://doi.org/10.3389/fpubh.2019.00170> (2019).
7. Eberhardt, K. A. et al. Ribavirin for the treatment of Lassa fever: A systematic review and meta-analysis. *Int. J. Infect. Dis.* **87**, 15–20 (2019).
8. Lukashevich, I. S., Paessler, S. & de la Torre, J. C. Lassa virus diversity and feasibility for universal prophylactic vaccine. *F1000Res* **8**, <https://doi.org/10.12688/f1000research.16989.1> (2019).
9. Purushotham, J., Lambe, T. & Gilbert, S. C. Vaccine platforms for the prevention of Lassa fever. *Immunol. Lett.* **215**, 1–11 (2019).
10. Mateo, M. et al. A single-shot Lassa vaccine induces long-term immunity and protects cynomolgus monkeys against heterologous strains. *Sci. Transl. Med.* **13**, eabf6348 (2021).
11. McCormick, J. B. et al. Lassa Fever. *N. Engl. J. Med.* **314**, 20–26 (1986).
12. Bell-Kareem, A. R. & Smither, A. R. Epidemiology of Lassa fever. in 1–23 (Springer, 2021). [https://doi.org/10.1007/82\\_2021\\_234](https://doi.org/10.1007/82_2021_234).
13. Nigeria Centre for Disease Control. <https://ncdc.gov.ng/diseases/sitreps/?cat=5&name=An%20update%20of%20Lassa%20fever%20outbreak%20in%20Nigeria>.
14. Manning, J. T., Forrester, N. & Paessler, S. Lassa virus isolates from Mali and the Ivory Coast represent an emerging fifth lineage. *Front. Microbiol.* **6**, <https://doi.org/10.3389/fmicb.2015.01037> (2015).
15. Dzotsi, E. K. et al. The first cases of Lassa fever in Ghana. *Ghana. Med. J.* **46**, 166–170 (2012).
16. Patassi, A. A. et al. Emergence of Lassa fever disease in northern Togo: Report of two cases in Oti District in 2016. *Case Rep. Infect. Dis.* **2017**, 8242313 (2017).
17. Yadouleton, A. et al. Lassa fever in Benin: Description of the 2014 and 2016 epidemics and genetic characterization of a new Lassa virus. *Emerg. Microbes Infect.* **9**, 1761–1770 (2020).
18. McCormick, J. B. & Fisher-Hoch, S. P. Lassa fever. *Curr. Top. Microbiol. Immunol.* **262**, 75–109 (2002).
19. Monath, T. P., Newhouse, V. F., Kemp, G. E., Setzer, H. W. & Cacciapuoti, A. Lassa virus isolation from *Mastomys natalensis* rodents during an epidemic in Sierra Leone. *Science* **185**, 263–265 (1974).
20. Stephenson, E. H., Larson, E. W. & Dominik, J. W. Effect of environmental factors on aerosol-induced Lassa virus infection. *J. Med. Virol.* **14**, 295–303 (1984).
21. Wozniak, D. M. et al. Inoculation route-dependent Lassa virus dissemination and shedding dynamics in the natural reservoir – *Mastomys natalensis*. *Emerg. Microbes Infect.* **10**, 2313–2325 (2021).
22. Ter Meulen, J. et al. Hunting of peridomestic rodents and consumption of their meat as possible risk factors for rodent-to-human transmission of Lassa virus in the Republic of Guinea. *Am. J. Trop. Med. Hyg.* **55**, 661–666 (1996).
23. Downs, I. L. et al. Natural history of aerosol induced Lassa fever in non-human primates. *Viruses* **12**, 593 (2020).



24. Lecompte, E. et al. *Mastomys natalensis* and Lassa Fever, West Africa. *Emerg. Infect. Dis.* **12**, 1971–1974 (2006).
25. Smither, A. R. & Bell-Kareem, A. R. Ecology of Lassa Virus. in 1–20 (Springer, 2021). [https://doi.org/10.1007/82\\_2020\\_231](https://doi.org/10.1007/82_2020_231).
26. Ogbu, O., Ajuluchukwu, E. & Uneke, C. J. Lassa fever in West African sub-region: An overview. *J. Vector Borne Dis.* **44**, 1–11 (2007).
27. Fichet-Calvet, E. et al. Fluctuation of abundance and Lassa virus prevalence in *Mastomys natalensis* in Guinea, West Africa. *Vector Borne Zoonotic Dis.* **7**, 119–128 (2007).
28. Fichet-Calvet, E., Becker-Ziaja, B., Koivogui, L. & Günther, S. Lassa serology in natural populations of rodents and horizontal transmission. *Vector Borne Zoonotic Dis.* **14**, 665–674 (2014).
29. Lo Iacono, G. et al. Using modelling to disentangle the relative contributions of zoonotic and anthroponotic transmission: The case of Lassa fever. *PLoS Negl. Trop. Dis.* **9**, e3398 (2015).
30. Siddle, K. J. et al. Genomic analysis of Lassa virus during an increase in cases in Nigeria in 2018. *N. Engl. J. Med.* **379**, 1745–1753 (2018).
31. Kafetzopoulou, L. E. et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* **363**, 74–77 (2019).
32. Andersen, K. G. et al. Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell* **162**, 738–750 (2015).
33. Lalis, A. & Wirth, T. Mice and men: An evolutionary history of Lassa fever. in *Biodiversity and Evolution* (eds. Grandcolas, P. & Maurel, M.-C.) 189–212, <https://doi.org/10.1016/B978-1-78548-277-9.50011-5> (Elsevier, 2018).
34. Mylne, A. Q. N. et al. Mapping the zoonotic niche of Lassa fever in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109**, 483–492 (2015).
35. Colangelo, P. et al. A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biol. J. Linn. Soc.* **108**, 901–916 (2013).
36. Gryseels, S. et al. When viruses don't go viral: The importance of host phylogeographic structure in the spatial spread of arenaviruses. *PLoS Path* **13**, e1006073 (2017).
37. Cuypers, L. N. et al. Three arenaviruses in three subspecific natal multimammate mouse taxa in Tanzania: Same host specificity, but different spatial genetic structure? *Virus Evol.* <https://doi.org/10.1093/ve/veaa039> (2020).
38. Vazeille, M., Gaborit, P., Mousson, L., Girod, R. & Failloux, A.-B. Competitive advantage of a dengue 4 virus when co-infecting the mosquito *Aedes aegypti* with a dengue 1 virus. *BMC Infect. Dis.* **16**, 318 (2016).
39. Chan, K. F. et al. Investigating viral interference between influenza A virus and human respiratory syncytial virus in a ferret model of infection. *J. Infect. Dis.* **218**, 406–417 (2018).
40. Meunier, D. Y., McCormick, J. B., Georges, A. J., Georges, M. C. & Gonzalez, J. P. Comparison of Lassa, Mobala, and Ippy virus reactions by immunofluorescence test. *Lancet* **1**, 873–874 (1985).
41. Howard, C. R. Antigenic diversity among the Arenaviruses. in *The Arenaviridae* (ed. Salvato, M. S.) 37–49, [https://doi.org/10.1007/978-1-4615-3028-2\\_3](https://doi.org/10.1007/978-1-4615-3028-2_3) (Springer US, 1993).
42. Bhattacharyya, S., Gesteland, P. H., Korgenski, K., Bjørnstad, O. N. & Adler, F. R. Cross-immunity between strains explains the dynamical pattern of paramyxoviruses. *Proc. Natl Acad. Sci. U. S. A.* **112**, 13396–13400 (2015).
43. Luis, A. D., Douglass, R. J., Mills, J. N. & Bjørnstad, O. N. Environmental fluctuations lead to predictability in Sin Nombre hantavirus outbreaks. *Ecology* **96**, 1691–1701 (2015).
44. Anderson, R. M., Jackson, H. C., May, R. M. & Smith, A. M. Population dynamics of fox rabies in Europe. *Nature* **289**, 765–771 (1981).
45. Tian, H. et al. Anthropogenically driven environmental changes shift the ecological dynamics of hemorrhagic fever with renal syndrome. *PLoS Pathog.* **13**, e1006198 (2017).
46. Redding, D. W., Moses, L. M., Cunningham, A. A., Wood, J. & Jones, K. E. Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever. *Methods Ecol. Evol.* **7**, 646–655 (2016).
47. Peterson, A. T., Moses, L. M. & Bausch, D. G. Mapping transmission risk of Lassa fever in West Africa: the importance of quality control, sampling bias, and error weighting. *PLoS One* **9**, e100711 (2014).
48. Fichet-Calvet, E. & Rogers, D. J. Risk maps of Lassa fever in West Africa. *PLoS Negl. Trop. Dis.* **3**, e388 (2009).
49. Basinski, A. J. et al. Bridging the gap: Using reservoir ecology and human serosurveys to estimate Lassa virus spillover in West Africa. *PLoS Comput. Biol.* **17**, e1008811 (2021).
50. Iacono, G. L. et al. A unified framework for the infection dynamics of zoonotic spillover and spread. *PLoS Negl. Trop. Dis.* **10**, e0004957 (2016).
51. Coumou, D. & Rahmstorf, S. A decade of weather extremes. *Nat. Clim. Change* **2**, 491–496 (2012).
52. Coumou, D., Robinson, A. & Rahmstorf, S. Global increase in record-breaking monthly-mean temperatures. *Clim. Change* **118**, 771–782 (2013).
53. Bathiany, S., Dakos, V., Scheffer, M. & Lenton, T. M. Climate models predict increasing temperature variability in poor countries. *Sci. Adv.* **4**, eaar5809 (2018).
54. Arneeth, A. Uncertain future for vegetation cover. *Nature* **524**, 44–45 (2015).
55. Brandt, M. et al. Human population growth offsets climate-driven increase in woody vegetation in sub-Saharan Africa. *Nat. Ecol. Evol.* **1**, 81 (2017).
56. Herrmann, S. M., Brandt, M., Rasmussen, K. & Fensholt, R. Accelerating land cover change in West Africa over four decades as population pressure increased. *Com. Earth Envir* **1**, 1–10 (2020).
57. Gibb, R., Moses, L. M., Redding, D. W. & Jones, K. E. Understanding the cryptic nature of Lassa fever in West Africa. *Pathog. Glob. Health* **111**, 276–288 (2017).
58. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
59. Frieler, K. et al. Assessing the impacts of 1.5 °C global warming—simulation protocol of the Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP2b). *Geosci. Model Dev.* **10**, 4321–4345 (2017).
60. Soberón, J. & Nakamura, M. Niches and distributional areas: Concepts, methods, and assumptions. *Proc. Natl Acad. Sci. U. S. A.* **106**, 19644–19650 (2009).
61. Moss, R. H. et al. The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756 (2010).
62. Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
63. Lukashevich, I. S. Generation of reassortants between African arenaviruses. *Virology* **188**, 600–605 (1992).
64. Vijaykrishna, D., Mukerji, R. & Smith, G. J. D. RNA virus reassortment: an evolutionary mechanism for host jumps and immune evasion. *PLoS Path* **11**, e1004902 (2015).
65. Whitmer, S. L. M. et al. New lineage of Lassa Virus, Togo, 2016. *Emerg. Infect. Dis.* **24**, 599 (2018).
66. Ehichioya, D. U. et al. Phylogeography of Lassa virus in Nigeria. *J. Virol.* **93**, e00929–19 (2019).
67. Dellicour, S., Rose, R., Faria, N. R., Lemey, P. & Pybus, O. G. SER-APHIM: studying environmental rasters and phylogenetically informed movements. *Bioinformatics* **32**, 3204–3206 (2016).

68. Dellicour, S. et al. Using viral gene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics. *Mol. Biol. Evol.* **34**, 2563–2571 (2017).
69. Dellicour, S. et al. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nat. Commun.* **11**, 5620 (2020).
70. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
71. Strahler, A. N. Quantitative analysis of watershed geomorphology. *Eos, Trans. Am. Geophys. Union* **38**, 913–920 (1957).
72. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
73. Ehichioya, D. U. et al. Current molecular epidemiology of Lassa virus in Nigeria. *J. Clin. Microbiol.* **49**, 1157 (2011).
74. Oloniniyi, O. K. et al. Genetic characterization of Lassa virus strains isolated from 2012 to 2016 in southeastern Nigeria. *PLoS Negl. Trop. Dis.* **12**, e0006971 (2018).
75. Olesen, J. E. et al. Uncertainties in projected impacts of climate change on European agriculture and terrestrial ecosystems based on scenarios from regional climate models. *Clim. Change* **81**, 123–143 (2007).
76. Simo Tchegnna, H. et al. Molecular characterization of a new highly divergent Mobala related arenavirus isolated from Praomys sp. rodents. *Sci. Rep.* **11**, 10188 (2021).
77. Olayemi, A. et al. New hosts of the Lassa virus. *Sci. Rep.* **6**, 25280 (2016).
78. Zaidi, M. B. et al. Competitive suppression of dengue virus replication occurs in chikungunya and dengue co-infected Mexican infants. *Parasit. Vectors* **11**, 378 (2018).
79. Olayemi, A. et al. Widespread arenavirus occurrence and seroprevalence in small mammals, Nigeria. *Parasit. Vectors* **11**, 416 (2018).
80. Nigeria Centre for Disease Control. <https://ncdc.gov.ng/diseases/sitreps/?cat=5&name=An%20update%20of%20Lassa%20fever%20outbreak%20in%20Nigeria>.
81. Norris, K. et al. Biodiversity in a forest agriculture mosaic: the changing face of west Africa rainforests. *Biol. Conserv.* **143**, 2341–2350 (2010).
82. Stocker, T. F. et al. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change.* (Cambridge University Press, Cambridge, United Kingdom and New York, 2013).
83. Buba, M. I. et al. Mortality among confirmed Lassa fever cases during the 2015–2016 outbreak in Nigeria. *Am. J. Public Health* **108**, 262–264 (2018).
84. Tobin, E. A. et al. Knowledge of secondary school children in Edo State on Lassa fever and its implications for prevention and control. *West. Afr. J. Med.* **34**, 101–107 (2015).
85. Saez, A. M. et al. Rodent control to fight Lassa fever: Evaluation and lessons learned from a 4-year study in Upper Guinea. *PLoS Negl. Trop. Dis.* **12**, e0006829 (2018).
86. Ejembi, J. et al. Contact tracing in Lassa fever outbreak response, an effective strategy for control? *Online J. Public Health Inf.* **11**, e378 (2019).
87. ECHO Flash List. <https://erccportal.jrc.ec.europa.eu/ECHO-Flash/ECHO-Flash-List/yy/2018/mm/2>.
88. Pigott, D. M. et al. Local, national, and regional viral haemorrhagic fever pandemic potential in Africa: a multistage analysis. *Lancet* **390**, 2662–2672 (2017).
89. Kraemer, M. U. G. et al. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nature Microbiology* <https://doi.org/10.1038/s41564-019-0376-y> (2019).
90. Elith, J. et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
91. Dhingra, M. S. et al. Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 2.3.4.4 viruses with spatial cross-validation. *eLife* **5**, e19571 (2016).
92. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
93. Elith, J. et al. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57 (2011).
94. Phillips, S. J. et al. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197 (2009).
95. Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**, 225–232 (2019).
96. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
97. Randin, C. F. et al. Are niche-based species distribution models transferable in space? *J. Biogeogr.* **33**, 1689–1703 (2006).
98. Lange, S. Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. *Earth Syst. Dyn.* **9**, 627–645 (2018).
99. Dunne, J. P. et al. GFDL's ESM2 global coupled climate-carbon earth system models. Part I: physical formulation and baseline simulation characteristics. *J. Clim.* **25**, 6646–6665 (2012).
100. Jones, C. D. et al. The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.* **4**, 543–570 (2011).
101. Dufresne, J.-L. et al. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dyn.* **40**, 2123–2165 (2013).
102. Watanabe, M. et al. Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *J. Clim.* **23**, 6312–6335 (2010).
103. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteor. Soc.* **93**, 485–498 (2012).
104. Hurtt, G. C. et al. Harmonization of global land-use change and management for the period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev.* 1–65 <https://doi.org/10.5194/gmd-2019-360> (2020)
105. Jones, B. & O'Neill, B. C. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environ. Res. Lett.* **11**, 084003 (2016).
106. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
107. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
108. Ayres, D. L. et al. BEAGLE 3: Improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Syst. Biol.* <https://doi.org/10.1093/sysbio/syz020> (2019).
109. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**, 57–86 (1986).
110. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
111. Laenen, L. et al. Spatio-temporal analysis of Nova virus, a divergent hantavirus circulating in the European mole in Belgium. *Mol. Ecol.* **25**, 5994–6008 (2016).
112. Dellicour, S. et al. Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: comparative study and analytical developments. *Heredity* **123**, 228–241 (2019).
113. Dellicour, S. et al. Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. *Nat. Commun.* **9**, 2222 (2018).



114. Dellicour, S. et al. Phylogeographic and phylodynamic approaches to epidemiological hypothesis testing. *bioRxiv* <https://doi.org/10.1101/788059> (2020).
115. Dellicour, S., Rose, R. & Pybus, O. G. Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinform* **17**, 1–12 (2016).
116. McRae, B. H. Isolation by resistance. *Evolution* **60**, 1551–1561 (2006).
117. Jacquot, M., Nomikou, K., Palmarini, M., Mertens, P. & Biek, R. Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference. *Proc. R. Soc. Lond. B* **284**, 20170919 (2017).
118. Gill, M. S. et al. Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
119. Karcher, M. D., Palacios, J. A., Bedford, T., Suchard, M. A. & Minin, V. N. Quantifying and mitigating the effect of preferential sampling on phylodynamic inference. *PLoS Comput. Biol.* **12**, e1004789 (2016).

## Acknowledgements

The authors thank David Pigott for sharing their environmental suitability predictions for *Mastomys natalensis*. The work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 725422 — ReservoirDOCS (M.A.S., P.L.), the European Union's Horizon 2020 project MOOD under grant agreement no. 874850 (P.L., S.Dellicour), the Wellcome Trust through project 206298/Z/17/Z—The Artic Network (M.A.S., P.L.), the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award numbers U01AI151812 (K.G.A.), R01AI153044 (P.L. and M.A.S.), and U19AI135995 (K.G.A.). S.Günther acknowledges support by the German Research Foundation (DFG, grants GU883/5-1 and GU883/5-2), and by the German Federal Ministry of Health through the WHO Collaborating Centre for Arboviruses and Hemorrhagic Fever Viruses at the Bernhard-Nocht-Institute for Tropical Medicine (agreement ZMV I1-2517WHO005) and the Global Health Protection Program (GHPP, agreements ZMV I1-2517GHP-704 and ZMV I1-2519GHP704). R.F.G. and D.S.G. acknowledge support from the Coalition for Epidemic Preparedness Innovation, the Wellcome Trust Foundation, and the European and Developing Countries Clinical Trials Partnership Programme. R.F.G. acknowledges support from the NIH (Grants R01AI132223, R01AI132244, U19AI142790, U54CA260581, U54HG007480, and OT2HL158260) and Gilead Sciences. B.V. was supported by a postdoctoral grant (12U7121N) of the Research Foundation—Flanders (*Fonds voor Wetenschappelijk Onderzoek*). S.Gryseels was supported by FED-tWIN OMEGA. P.L. acknowledges support from the Research Foundation — Flanders (*Fonds voor Wetenschappelijk Onderzoek — Vlaanderen*; grant no. G066215N, GOD5117N, and GOB9317N). S.Dellicour acknowledges support from the *Fonds National de la Recherche Scientifique* (F.R.S.-FNRS, Belgium; grant no. F.4515.22) and from the Research Foundation — Flanders (*Fonds voor Wetenschappelijk Onderzoek — Vlaanderen*; grant no. G098321N). The content is solely the responsibility of the authors and does not necessarily

represent the official views of the National Institutes of Health. Computational resources have been provided by the *Consortium des Équipements de Calcul Intensif* (CÉCI), funded by the *Fonds de la Recherche Scientifique de Belgique* (F.R.S.-FNRS) under Grant no. 2.5020.11 and by the Walloon Region.

## Author contributions

Study design: R.K., W.T., G.D., M.A.S., P.L., K.G.A., and S.Dellicour. Sequencing: L.E.K., S.Duraffour, and S.Günther. Data collection and curation: R.K., L.E.K., W.T., S.Gryseels, A.K., K.G., M.Momoh, J.D.S., A.G., F.A, D.S.G., S.O., E.O.E., R.F.G., A.R.S., M.G.P, M.McGraw, L.D.H, and S.Dellicour. Data analysis: R.K., W.T., G.D., K.G., M.Z., B.V., and S.Dellicour. Writing - original draft preparation: R.K., W.T., P.L., K.G.A., and S.Dellicour. Writing - review and editing: R.K., W.T., B.V., K.G., M.Z., M.A.S., P.L., K.G.A., and S.Dellicour. Supervision: P.L., K.G.A., and S.Dellicour.

## Competing interests

R.F.G. is a co-founder of Zalgen Labs, a biotechnology company developing countermeasures for emerging viruses. The other authors declare no competing interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33112-3>.

**Correspondence** and requests for materials should be addressed to Raphaëlle Klitting or Simon Dellicour.

**Peer review information** *Nature Communications* thanks James Wood and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

<sup>1</sup>Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>2</sup>Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory for Clinical and Epidemiological Virology, KU Leuven - University of Leuven, Leuven, Belgium. <sup>3</sup>Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. <sup>4</sup>Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, Belgium. <sup>5</sup>Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania. <sup>6</sup>Evolutionary Ecology group, Department of Biology, University of Antwerp, 2610 Antwerp, Belgium. <sup>7</sup>Vertebrate group, Directorate Taxonomy and Phylogeny, Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium. <sup>8</sup>Eastern Technical University of Sierra Leone, Kenema, Sierra Leone. <sup>9</sup>Viral Hemorrhagic Fever Program, Kenema Government Hospital, Ministry of Health and Sanitation, Kenema, Sierra Leone. <sup>10</sup>College of Medicine and Allied Health Sciences, University of Sierra Leone, Kenema, Sierra Leone.

<sup>11</sup>Irrua Specialist Teaching Hospital, Irrua, Nigeria. <sup>12</sup>Faculty of Clinical Sciences, College of Medicine, Ambrose Alli University, Ekpoma, Nigeria. <sup>13</sup>Department of Microbiology and Immunology, Tulane University, School of Medicine, New Orleans, LA 70112, USA. <sup>14</sup>Zalgen Labs, LCC, Frederick, MD 21703, USA. <sup>15</sup>Global Virus Network (GVN), Baltimore, MD 21201, USA. <sup>16</sup>Department of Integrative, Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>17</sup>German Center for Infection Research (DZIF), Partner site Hamburg–Lübeck–Borstel–Riems, Hamburg, Germany. <sup>18</sup>Department of Biostatistics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. <sup>19</sup>Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA. <sup>20</sup>Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. <sup>21</sup>Scripps Research Translational Institute, La Jolla, CA 92037, USA. <sup>22</sup>Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, CP160/12 50, av. FD Roosevelt, 1050 Bruxelles, Belgium. ✉ e-mail: [rkitting@scripps.edu](mailto:rkitting@scripps.edu); [simon.dellicour@ulb.be](mailto:simon.dellicour@ulb.be)