VIRUS EVOLUTION

Virus Evolution, 2025, **11(1)**, veaf017
DOI: https://doi.org/10.1093/ve/veaf017
Advance Access Publication 23 September 2025
Research Article

Tracking down the origin and subsequent spread of SARS-CoV-2 lineage B.1.619

Nena Bollen[1,*], Samuel L. Hong[1], Barney I. Potter[1], Reto Lienhard[2], Marie-Lise Tritten[2], Nicolas Sierro[3], Emmanuel Guedj[3], Rémi Dulize[3], David Bornand[3], Mehdi Auberson[3], Maxime Berthouzoz[3], Pauline Duvoisin[3], Nikolai V. Ivanov[3], Manuel C. Peitsch[3], Verity Hill[4], COVID-19 Genomics Belgium Consortium[†], Veerle Matheeussen[5], Sébastien Bontems[6], Bruno Verhasselt[7], Jonathan Degosserie[8], Luc Waumans[9], Guillaume Bayon-Vicente[10], Marijke Reynders[11], Lien Cattoir[12], Valentin Coste[13], Hanne Valgaeren[14], Johan Van Weyenbergh[1], Lize Cuypers[15], Emmanuel André[15,16], Keith Durkin[17], Piet Maes[1], Kamran Khan[18,19,20], Carmen Huber[18], Marc A. Suchard[21,22,23], Martin Maidadi Foudi[24], Célestin Godwe[24], Moise Henri Moumbeket Yifomnjou[25], Messanga Landry[25], Richard Njouom[25], Placide Mbala Kingebeni[26,27], Paul Oluniyi[28,29], Idowu B. Olawoye[28,29], Christian Happi[28,29], Ahidjo Ayouba[30], Martine Peeters[30], Sylvie Behillil[31], Etienne Simon-Lorière[31], Martin Hölzer[32], Simon Dellicour[1,33], Gytis Dudas[34], Guy Baele[1]

[1]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium
[2]ADMED Microbiologie, Boucle de Cydalise 16, La Chaux-de-Fonds 2300, Switzerland
[3]PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, Neuchâtel 2000, Switzerland
[4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom
[5]Department of Microbiology, University Hospital Antwerp, Edegem and Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Wilrijk, Belgium
[6]Department of Clinical Microbiology, University of Liège, Liège 4000, Belgium
[7]Department of Diagnostic Sciences, Ghent University Hospital, Ghent University, Ghent 9000, Belgium
[8]Department of Laboratory Medicine, Université Catholique de Louvain, CHU UCL Namur, Yvoir 5530, Belgium
[9]Clinical Laboratory, Jessa Hospital, Hasselt, Belgium
[10]Laboratory of Proteomics and Microbiology, Research Institute for Biosciences, University of Mons, Mons, Belgium
[11]Division of Medical Microbiology, Department of Laboratory Medicine, AZ Sint-Jan Brugge-Oostende AV, Brugge, Belgium
[12]Department of Laboratory Medicine, OLV Hospital, Aalst, Belgium
[13]Metabolism and Nutrition Research Group, Louvain Drug Research Institute, Université Catholique De Louvain, UCLouvain, Brussels, Belgium
[14]Clinical Laboratory, Imelda Hospital, Bonheiden 2820, Belgium
[15]Department of Laboratory Medicine, UZ Leuven Hospital, Herestraat 49, Leuven 3000, Belgium
[16]Department of Microbiology, Immunology and Transplantation, Laboratory of Clinical Bacteriology and Mycology, Rega Institute, KU Leuven, Leuven 3000, Belgium
[17]Department of Human Genetics, CHU Liège, and Medical Genomics, GIGA Research Center, University of Liège, Liège, Belgium
[18]BlueDot, Toronto, ON M5J 1A7, Canada
[19]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON M5B 1A6, Canada
[20]Section of Infectious Diseases, Department of Medicine, University of Toronto, Toronto, ON M5S 3H2, Canada
[21]Department of Biomathematics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States
[22]Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States
[23]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States
[24]Centre de Recherche en Maladies Emergentes et Re-emergentes (CREMER), Institut de Recherche M'dicales et d'étude des Plantes Médicinales (IMPM), Ministère de la Recherche Scientifique et de l'Innovation (MINRESI), Yaounde, Cameroon
[25]Virology Department, Centre Pasteur du Cameroun, Yaounde, Cameroon
[26]Institut National de Recherche Biomédicale, Ministère de la Santé Publique, Kinshasa-Gombe B.P. 1197, Democratic Republic of the Congo (DRC)
[27]L'Hôpital Général de Référence de Kole, Kole, Democratic Republic of the Congo
[28]Department of Biological Sciences, Redeemer's University, Ede, Osun 232101, Nigeria
[29]African Centre of Excellence for Genomics of Infectious Diseases, Ede, Osun 232101, Nigeria
[30]TransVIHMI, Université de Montpellier, IRD, INSERM, 911 Avenue Agropolis, Montpellier cedex 34394, France
[31]Evolutionary Genomics of RNA Viruses Group, Institut Pasteur, Paris, France
[32]Bioinformatics and Systems Biology, Methodology and Research Infrastructure, Robert Koch Institute, Berlin, Germany
[33]Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Bruxelles, Belgium
[34]Institute of Biotechnology, Life Sciences Centre, Vilnius University, Vilnius, Lithuania

[*]Corresponding author. Herestraat 49 - box 1030B-3000 Leuven. E-mail: nena.bollen@kuleuven.be
[†]Membership of the COVID-19 Genomics Belgium consortium is listed in the Acknowledgments.

Abstract

Since late 2020, the emergence of variants of concern (VOCs) of SARS-CoV-2 has been of concern to public health, researchers and policymakers. Mutations in the SARS-CoV-2 genome—for which clear evidence is available indicating a significant impact on transmissibility, severity and/or immunity—illustrate the importance of genomic surveillance and monitoring the evolution and geographic spread of novel lineages. Lineage B.1.619 was first detected in Switzerland in January 2021, in international travellers returning from Cameroon. This lineage was subsequently also detected in Rwanda, Belgium, Cameroon, France, and many other countries and is characterised by spike protein amino acid mutations N440K and E484K in the receptor binding domain, which are associated with immune escape and higher infectiousness. In this study, we perform a phylogeographic analysis to track the geographic origin and subsequent dispersal of SARS-CoV-2 lineage B.1.619. We employ a recently developed travel history-aware phylogeographic model, enabling us to incorporate genomic sequences with associated travel information. We estimate that B.1.619 most likely originated in Cameroon, in November 2020. We estimate the influence of the number of air-traffic passengers on the dispersal of B.1.619 but find no significant effect, illustrative of the complex dispersal patterns of SARS-CoV-2 lineages. Finally, we examine the metadata associated with infected Belgian patients and report a wide range of symptoms and medical interventions.

Keywords: SARS-CoV-2; COVID-19; B.1.619; phylogenetics; phylogeography; GLM; air traffic; Bayesian inference; Markov chain Monte Carlo

Introduction

More than 3 years since its discovery, the SARS-CoV-2 pandemic is responsible for 630 million known infections and at least 6.5 million deaths worldwide (WHO, 2022). During this period, unprecedented research efforts were made to cope with the pandemic, while at the same time health systems were under high pressure. Globally, significant efforts to track the spread and evolution of the virus continue to be made, with nearly 14 million SARS-CoV-2 genomes currently available on the Global Initiative on Sharing Avian Influence Data (GISAID) (Shu and McCauley, 2017).

The discovery of several new variants of concern (VOCs) at the end of 2020 caused additional worry for public health officials. The first notable example was lineage B.1.1.7 (also denominated as the Alpha variant by the World Health Organisation; WHO), first discovered in the United Kingdom in late 2020 (Davies et al., 2021; Funk et al., 2021). Another VOC of note is the B.1.617.2 lineage (also known as the Delta variant per WHO designation), first detected in India, which was characterised by higher transmissibility and some resistance to vaccine-induced immunity (Funk et al., 2021; Planas et al., 2021). Delta became the dominant lineage worldwide in summer 2021. In November 2021, the B.1.1.529 lineage—discovered in southern Africa and denominated Omicron by the WHO—caused an unprecedented global surge in cases and its descendant lineages continue to circulate to this day. Like other VOCs, this lineage is marked with significant antigenic advance, leading it to rapidly displace Delta as the dominant strain worldwide (Planas et al., 2021). In contrast to Alpha and Delta, Omicron was found to actually be less virulent (Sigal, 2022).

Apart from the designation and tracking of VOCs, it might be beneficial to also label some mutations as 'mutations of concern' and to intensify their molecular surveillance. Certain mutations found in these VOCs are of particular interest, especially those occurring in the spike protein. For example, mutations K417N, E484K, N501Y, H655Y, T478K and D614G are all found in at least two of the VOCs mentioned above. The current Omicron diversity shows a high level of re-occurring mutations caused by convergent evolution, with several associated to increased resistance to antibodies (Hoffmann et al., 2022; Jung et al., 2022). In this study, we focus on SARS-CoV-2 lineage B.1.619.

Although this lineage has not spread widely as the VOCs and is probably now extinct, we believe it deserves attention because it has a number of concerning mutations (such as E484K). Tracking such mutations of concern can help us understand how they interact and evolve, and highlight lineages likely to pose a danger for detailed inquiry. When the wave of infection attributable to B.1.619 disappeared in the summer of 2021 (Figure 1), new reports were published on the emergence of a pango sub-lineage called B.1.619.1 circulating primarily in South Korea which has an additional mutation in ORF1ab (K3929R) (Park et al., 2022; Ruis, 2022). Analysing the specific aspects of this sub-lineage is, however, beyond the scope of this study.

Lineage B.1.619 was first sampled in Switzerland in travellers who had just returned from Cameroon around the 16th of January 2021 (see Figure 1). For this reason, it is probable that its immediate origins for the European outbreak lie in Cameroon. However, genomic surveillance of SARS-CoV-2 in African countries remains at a lower level than in other countries with higher incidence rates (Brito et al., 2022), and researchers have pointed out undersampling and under-reporting as a possible issue (Adebisi et al., 2020; Rice et al., 2021). As noted by Rice et al.: 'variation in reporting between countries and some seroprevalence surveys that suggested high rates of local infection make it unclear if the relatively few reported cases and deaths to date indicate a generally reduced epidemic potential' (Rice et al., 2021). Such undersampling is a problem, especially in the case of phylogeographic analyses, where early-epidemic sequences are crucial to estimate the dispersal history of a virus.

Therefore, under-sampling is a potential limitation to our study, since the goal was to perform a phylogeographic reconstruction of the spread of B.1.619, estimating both its origin location and time. Thankfully, we have valuable information at our disposal. As mentioned before, the lineage was first sampled in travellers coming from Cameroon. In contrast to local case reporting, travel surveillance relies on diagnosing patients that have acquired infections while traveling outside the country of diagnosis (Grubaugh et al., 2019). In this study, we use a recently developed phylogeographic model, which explicitly incorporates travel origin and destination (Lemey et al., 2020). Models based on infectious disease surveillance of international travellers have proven effective in detecting pathogens circulating in resource-limited areas (Harvey et al., 2013; Leder et al., 2013; Wilder-Smith et al., 2014; Hamer et al., 2016; Grubaugh et al., 2019; Dudas et al., 2021; Kaleta et al., 2022).

Our results provide estimates of the geographic and temporal spread of lineage B.1.619, taking into account individual travel histories associated with nine samples. We also analysed the potential effect of various predictors on the spread of SARS-CoV-2 lineage B.1.619, namely geographical distance, border sharing and air-passenger volume. As such, our work provides a comprehensive phylogeographic analysis of the dispersal history and drivers of spread for lineage B.1.619, as well as an in-depth look at the mutational profile of the lineage. We conclude this study by examining the metadata associated with Belgian patients who

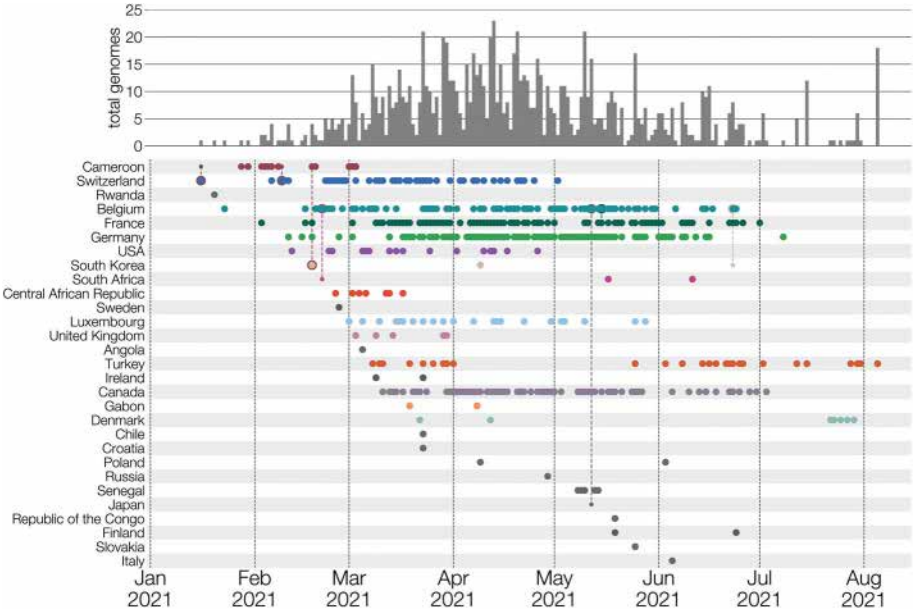

Figure 1. Known locations and travel history of B.1.619 cases. Collection dates of B.1.619 genomes are shown for each country (rows). Genomes from travellers are outlined with colour indicating travel of origin (*e.g.* dark red for Cameroon) and connected to a smaller dot indicating which country's diversity is being sampled at travel destination. Bars at the top indicate the number of genomes of B.1.619 available for a given date across all countries. Of note, 46 B.1.619 genomes from Canada were not included due to imprecise collection dates. On January 20, 2021, a single B.1.619 case was found in Rwanda; the travel case on May 1, 2021, was entering Japan from South Korea.

were infected with B.1.619 and report a wide range of symptoms and medical interventions.

Results

Notable mutations in the B.1.619 spike protein

As with any lineage, B.1.619 is characterised by the presence of several mutations, insertions and deletions in its genome, as can be seen in Figure 2. Compared to the SARS-CoV-2 reference sequence (NC_045512), there are 21 new mutations and two deletions that characterise B.1.619. It shares a number of mutations that have previously been observed in various VOCs and variants of interest (VOIs). Of major concern is the E484K mutation, which is present in both the Beta and Gamma variants (while in Omicron a similar mutation E484A is present), as well as a number of B.1.1.7 sequences (Frampton et al., 2021). This mutation has been associated with immune escape and higher fitness in immunised populations, with some researchers stating that: 'the presence of the E484K mutation by itself should be enough to qualify a variant for VOI status' (Boehm et al., 2021; Jaspe et al., 2021; Tandel et al., 2021). *In vitro* studies have shown an association of E484K with lower antibody binding, both in the case of monoclonal antibodies and in polyclonal antibodies from convalescent plasma (Jaspe et al., 2021; Greaney et al., 2021; Harvey et al., 2021; Sikora et al., 2021; Tandel et al., 2021) as well as in sera from vaccinated individuals (Jangra et al., 2021). Greaney and colleagues found that neutralisation by some sera was reduced up to 10-fold compared to the initial strain (Greaney et al., 2021). *In vivo*, E484K has been shown to be linked to re-infection in a Brazilian study (Nonaka et al., 2021).

Another mutation of interest present in B.1.619 is the N440K spike protein mutation, also seen in Omicron lineages. Tandel et al. (2021) found that this mutation is able to generate a larger amount of viruses in a shorter time, compared to previous variants (Tandel et al., 2021), which could partly help to explain Omicron's rapid spread. Furthermore, Augusto and colleagues found that this mutation showed a higher affinity for binding to the host receptor ACE2, which could correlate with higher infectiousness (Augusto et al., 2021). The mutation had been observed to be increasing in prevalence in some countries in the summer of 2021, including India and to a lesser extent the USA and Germany (Tandel et al., 2021). However, with the Omicron variant becoming dominant in the winter of 2021, this mutation is now widespread globally.

A third spike mutation worth mentioning is the H146Y mutation. McCallum et al. (2021) found that H146Y reduced binding of various antibodies (S2M28, S2X28, and in particular of 4A8), indicating a higher chance of immune escape. Together, E484K, N440K and H146Y paint a worrying picture, which makes the B.1.619 variant worth investigating.

A handful of SARS-CoV-2 genomes on GISAID from 2020 designated as lineage B.1.619 by Pangolin are missing key mutations (S:E484K, S:T1027I) that were present in the predecessor lineage of lineages B.1.619 and B.1.620, and thus were probably classified incorrectly due to sharing convergent mutations with B.1.619 (*e.g.* NSP12:P323L), these were not included in the analysis.

B.1.619 origin and spread

An important goal in our study was to estimate the global origin and spread of the B.1.619 lineage using Bayesian phylogeographic inference. The earliest known cases of B.1.619 were detected in travellers returning from Cameroon, which suggests that the African continent could be the origin of this lineage. Our analysis of the available genomes and accompanying travel information provides further evidence that this is likely the case. In addition to showing the number of B.1.619 genomes through time, Figure 1 illustrates the travel cases that were sampled.

Our first phylogeographic analysis (at the continent level) utilises both 723 B.1.619 sequences and over 543 'background' SARS-CoV-2 sequences to provide a wider context for the B.1.619

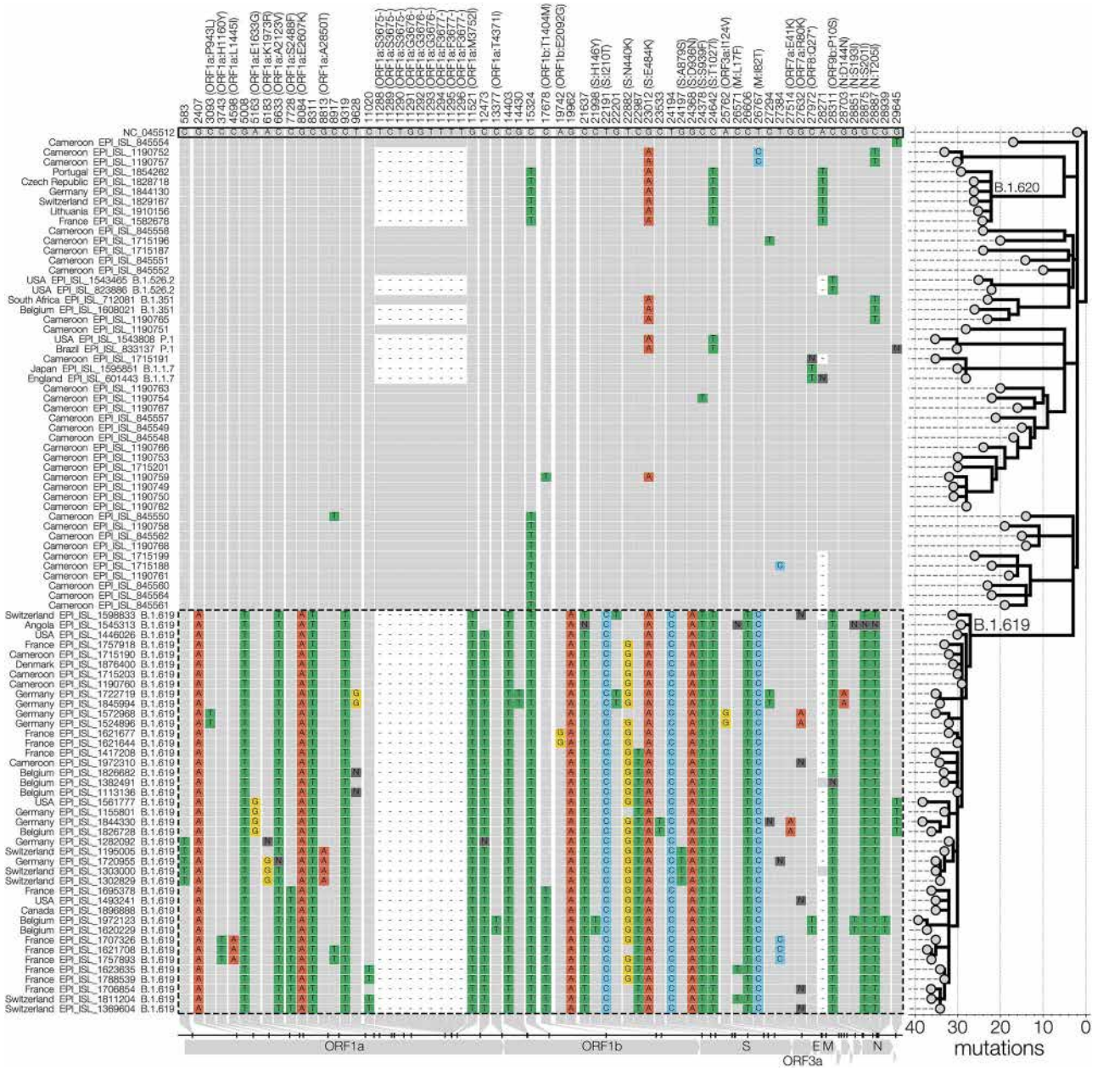

Figure 2. Lineage-defining SNPs of lineage B.1.619. Only SNPs that differentiate B.1.619 (genomes within the dashed line box) from the reference and are shared by at least two B.1.619 genomes are shown in the condensed SNP alignment. Representative genomes from lineage B.1.620 are included for comparison. Sites identical to the reference (GenBank accession NC_045512) are shown in grey, changes from the reference are indicated and coloured by nucleotide (green for thymine, red for adenosine, blue for cytosine, yellow for guanine, dark grey for ambiguities, black for gaps). If a mutation results in an amino acid change, the column label indicates the gene, reference amino acid, amino acid site, and amino acid change in brackets. The phylogeny (branch lengths number of mutations) on the right shows the relationships between depicted genomes and was rooted on the reference sequence.

epidemic. We also include a reference sequence (LR757998.1), which can be clearly seen forming a separate clade at the root. The results of this analysis can be seen in Figure 3A, where the B.1.619 clade is indicated with an asterisk. The estimated origin of this pandemic is estimated to be in Asia, this is explained by the inclusion of an early reference sequence from Wuhan. The mean evolutionary rate is 7.52E-4 (95 per cent highest posterior density (HPD): [6.89E-4; 8.16E-4]) substitutions per site per year and the estimated time to the most recent common ancestor (tMRCA) is 2019-12-07 (95 per cent HPD: [2019-10-17 ; 2019-12-26]). This analysis places the origin of the B.1.619 clade in Africa, with good posterior support (0.79). The backbone of the clade is also located in Africa. This is not unexpected, since we selected context sequences based on the African Nextstrain build (downloaded on June 25, 2021) (Hadfield et al., 2018). We estimated the highest number of Markov jumps from Africa into Asia and Europe (2.88, 95 per cent HPD: [1.05 ; 4.88] and 3.4, 95 per cent HPD: [1.39; 5.75], respectively). Markov jumps were also estimated to have occurred from Africa into North America (1.12, 95 per cent HPD: [0.35; 2.05]) and from Europe back into Africa (0.99, 95 per cent HPD: [0.15; 2.00]).

Our second phylogeographic analysis focuses only on the B.1.619 clade and uses more precise discrete locations at the country level. The country-level analysis reaches a similar

conclusion as the continent-level analysis, with the origin being estimated in Cameroon (see Figure 3B). The Markov jump analysis for the country level analysis shows multiple introduction events into Europe (mainly Germany, but also into France and Belgium) originating from Africa (see Figure 4). The analysis also shows multiple Markov jumps between European countries. The Markov jump plots showing the results for different months can be seen in Supplementary Figure 9, which shows that, although most jumps were not from Cameroon, the first detected jumps were.

We examined if the number of flights between two locations is correlated with the migration rates between locations using a GLM extension to the Bayesian phylogeographic analysis (Lemey et al., 2014). We also analysed the possible effect of distances between countries and whether two countries share a land border. The results of the GLM analysis can be seen in Figure 5, the results of the epoch-GLM analysis can be seen in Supplementary Figure 10. We find no significant role for passenger volumes between countries on the migration rates. On the other hand, there is relatively strong negative support for the role of distance, and very low positive support for the neighbour effect. In other words, for pairs of countries that were located far from each other, we had a lower probability of observing a migration event. For neighbouring countries (which share a land border), there might be a (non-significant) higher chance of a migration event taking place. Given the fact that travel restrictions might have had a significant effect on flights during our study period, we also examined the effect of passenger volume on a monthly basis. We ran an epoch-GLM model—assuming a GLM for each month—using the corresponding passenger volume matrices. As in the previous analysis, there was no significant effect of passenger volumes on migration during any month. In this analysis, neither distance nor border sharing were significant, with the exception of April 2021 where distance showed a significant negative effect (see Supplementary Figure 10).

B.1.619 in Belgium

Since we had access to detailed location information for the Belgian B.1.619 sequences, we were able to run a continuous phylogeographic analysis focused on Belgium. The results of this analysis allow us to estimate local dispersal within Belgium. Most cases of B.1.619 in Belgium are geographically spread out, with no apparent large clusters. As can be seen in Figure 6, there appears to have been an early outbreak of B.1.619 in Charleroi, and another one to the south of the city of Antwerp. Later, there are some small local clusters in various locations both in Flanders and in Wallonia. Note also that most samples were dated towards the end of our study period. The results of this analysis suggest that B.1.619 was sporadic in Belgium, not really gaining a foothold or causing any large outbreaks. The study period (summer of 2021) in Belgium was a transitional period in terms of lineage dominance, with the Alpha wave quickly ceding to Delta during the months of April to July (Cuypers et al., 2022).

For 129 Belgian B.1.619 sequences, we obtained information concerning the vaccination status of the patient and their location within Belgium, their age and gender as well as the health outcomes after infection. Although these data concerned only Belgium, they can still shed some light on the effects of B.1.619 on health outcomes. In this dataset, we identified two cases of post-vaccination infection, one after one dose and another after two doses of the vaccine. We have vaccination data on 28 patients, including the vaccination date for 9 individuals. This represents a vaccination coverage of 21 per cent for our sample. At the beginning of July (the end of our study period), the cumulative percentage of people having received one dose in Belgium was around 65 per cent, a 20 per cent coverage was achieved in mid-April, implying a slight under-representation of vaccination coverage in our sample.

As can be seen in Figure 7, more than half of the patients in the sample did not have any significant symptoms. The most common symptom was fatigue, followed by headache and fever. Few people experienced rhinitis, which is typical for SARS-CoV-2 infections. Regarding the health outcomes in this sample, most people did not require hospitalisation, and only five required stays in the intensive care unit (ICU). Note that for nearly all of the samples in this dataset, we do not know if the patients were infected pre- or post-vaccination, so it is not possible to make any conclusions about the percentage of patients experiencing symptoms in the vaccinated versus unvaccinated group. The average age in our sample for people who were hospitalised is 55, with the youngest hospitalised patient being 17 and the oldest being 82. Note that one person had no reported symptoms, but still required hospitalisation, which implies that symptoms were not reported for this patient. This was also a case of post-vaccination infection, occurring almost exactly a month after the first dose.

Discussion

Africa as the probable origin of B.1.619

As can be seen in Figure 3B, the backbone of the B.1.619 clade is mostly located within Cameroon. Our analysis showed initial transition events from Africa into various countries, followed later by more 'local' transition events between European nations, as well as local clusters indicating community transmissions. Based on available data and our analysis, it seems likely that B.1.619 indeed originated in Africa, and more specifically in Cameroon.

It should be noted, however, that this conclusion, like any analysis, is based only on the publicly available data. This might seem an evident conclusion but has important consequences, as it implies we can not infer an origin that was not included in our sample set. New sequence data could in theory drastically alter the conclusions in this paper. For this reason, we do not exclude the possibility that the true origin of the B.1.619 clade was in another African country, or maybe another continent altogether.

Not only a lack of data, but also biased sampling could alter our results. The number of available sequences at a given location is not always proportional to the number of cases at that location (Brito et al., 2022). Specifically in the case of discrete phylogeographic models such as the one employed in this study, highly sampled locations tend to skew the results with regards to the epidemic origin. In this case, despite having only 20 Cameroonian sequences in the analysis, it was still selected as the origin location for the B.1.619 epidemic. However, not many sequences were included from other African countries. In our phylogeny, the Cameroonian sequences are located near the root, without forming a clearly delineated clade. It would be hasty to conclude that B.1.619 did not manage to initiate community transmissions in any (other) African countries, or that there were no introductions into these countries. It should not be assumed that the variant is contained only to the countries in this analysis. A study by Salyer et al. (2021) showed that 9 of the 55 African Union (AU) member nations accounted for more than 80 per cent of reported cases, illustrating the need for more widespread and consistent surveillance networks.

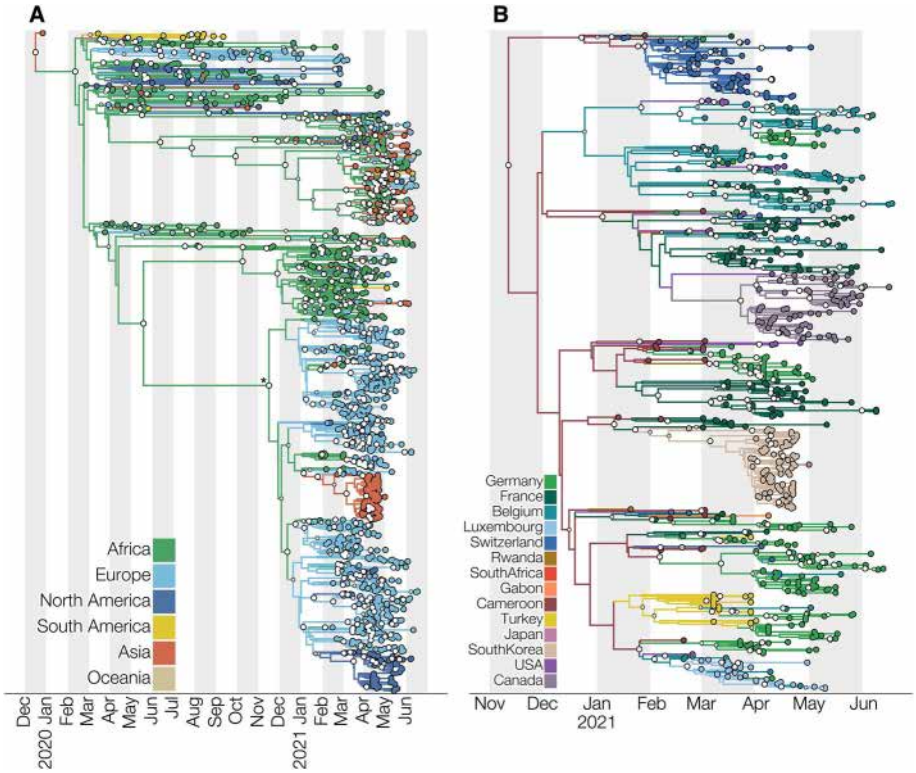

Figure 3. Maximum clade credibility trees of lineage B.1.619 coloured by reconstructed location using the latest available data as of June 2021. (A) Global phylogeny of SARS-CoV-2 genomes with branches coloured by inferred continent from a Bayesian phylogeographic analysis that makes use of individual travel histories. An asterisk indicates the root for the B.1.619 clade. Lineage B.1.619 is outlined and a horizontal bar shows the posterior probability of its common ancestor existing in a given continent. Africa is reconstructed as the most likely location (posterior probability 0.995) where B.1.619 originated. (B) Phylogeny of lineage B.1.619 with branches coloured by inferred country from a Bayesian phylogeographic analysis that makes use of travel histories. In this analysis Cameroon and Central African Republic (CAR) are reconstructed as the most likely locations (with posterior probabilities of 0.81 and 0.17, respectively) of the common ancestor of lineage B.1.619. Larger white dots at nodes indicate nodes with posterior probability of at least 95 per cent, while smaller grey circles indicate nodes with posterior probability of at least 50 per cent.

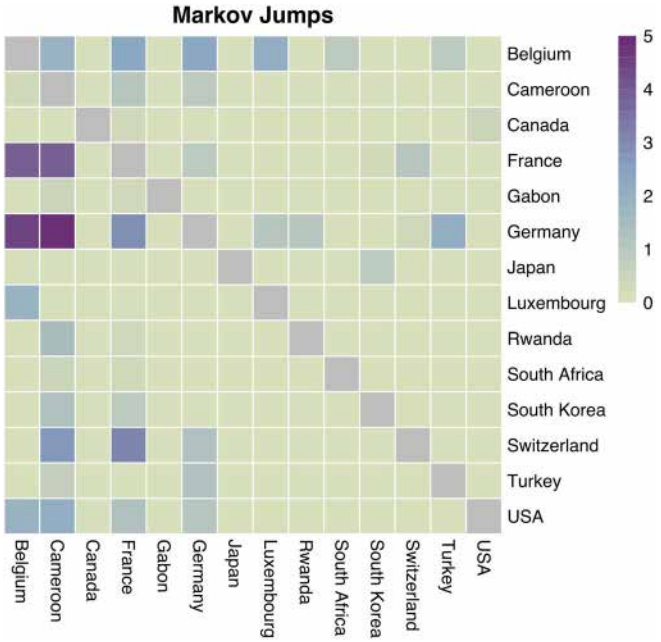

Figure 4. Markov jumps plot showing the number of B.1.619 transition events between countries. Countries of origin are shown in the columns, destination countries are shown in the rows. We see most jumps from both Cameroon and Belgium into Germany and France. Cameroon also seeds infections in other countries such as the USA, Switzerland, Rwanda and Belgium (among others).

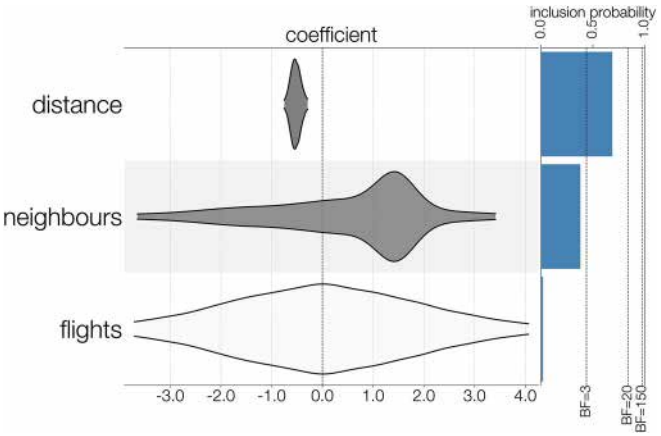

Figure 5. Identified predictors for the spread of B.1.619 We see that the only variable with a significant negative effect on the spread of B.1.619 is 'distance'. This means we estimate a reduced probability of seeing a transition event between country pairs with a larger distance between them. We see no effect at all for the number of flights between country pairs.

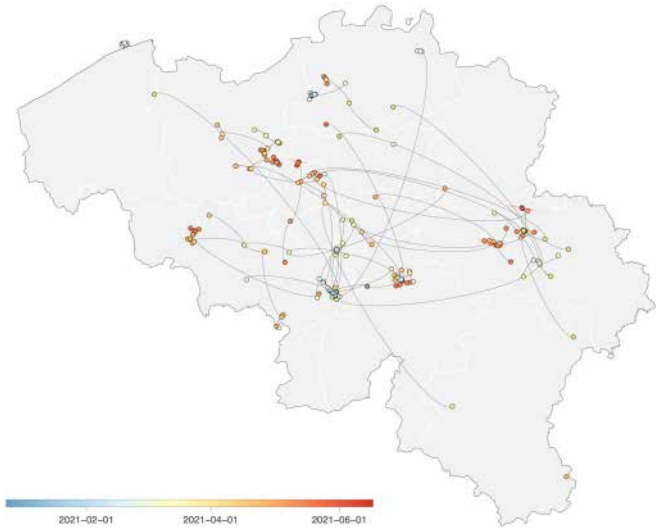

Figure 6. Spatially-explicit phylogeographic reconstruction of the dispersal history of the B.1.619 lineage in Belgium. Nodes are coloured from blue (the time of the most recent common ancestor, tMRCA) to red (most recent sampling time). The dispersal of B.1.619 within Belgium was characterized by both.

International travel and its effect on SARS-CoV-2 transmission

Our analysis did not show any significant effect of air-travel volume on the transmission of B.1.619 between countries, even when accounting for the variability of air traffic in different months. Other studies have shown that air traffic can have a positive significant effect on SARS-CoV-2 transmissions. For example the work by (Lau et al., 2020), found a strong linear correlation between the number of COVID-19 cases and passenger volume both within China (domestic flights) and on a global level (international flights) (Lau et al., 2020). Another study, by Sokadjo and Atchadé (2020) found a similar association. However, it is possible that both variables are associated with a higher population density, and that there is not necessarily a causal effect. These studies do not take into account the direction of flights (from one location to another), simply their passenger volume, nor do they measure the number of estimated introduction events. Intuitively, it seems that a higher number of flights would increase the probability of an introduction event. The more flights, the more separate opportunities that a successful transmission chain starts. However, the intensity of the epidemic at the potential donor site must also be considered. Higher case numbers will also influence the probability of an introduction event happening into another country, and is a factor often overlooked in cases where travel bans have been issued in the past (often based on the detection of only a few cases) (Grubaugh et al., 2019). Furthermore, only a few transition events are often enough to trigger a chain of local transmissions. Other studies have identified specific individual introduction events as a consequence of international air travel by incorporating known travel cases into a phylogenetic analysis (as was done in our own analysis) (Alteri et al., 2021; Dudas et al., 2021). In the case of the 2014 Ebola epidemic, it is believed that an infected individual travelling internationally by land, might have been at the root of many new clusters of infections (Yang et al., 2015). So it is not necessarily always the sheer number of flights that influences the occurrence of an epidemic.

Understanding the effect of air travel on the spread of SARS-CoV-2 is of great importance, considering that a commonly employed strategy has involved border closures or flight restrictions. During 2020 and 2021, many countries completely closed their borders during certain periods and implemented

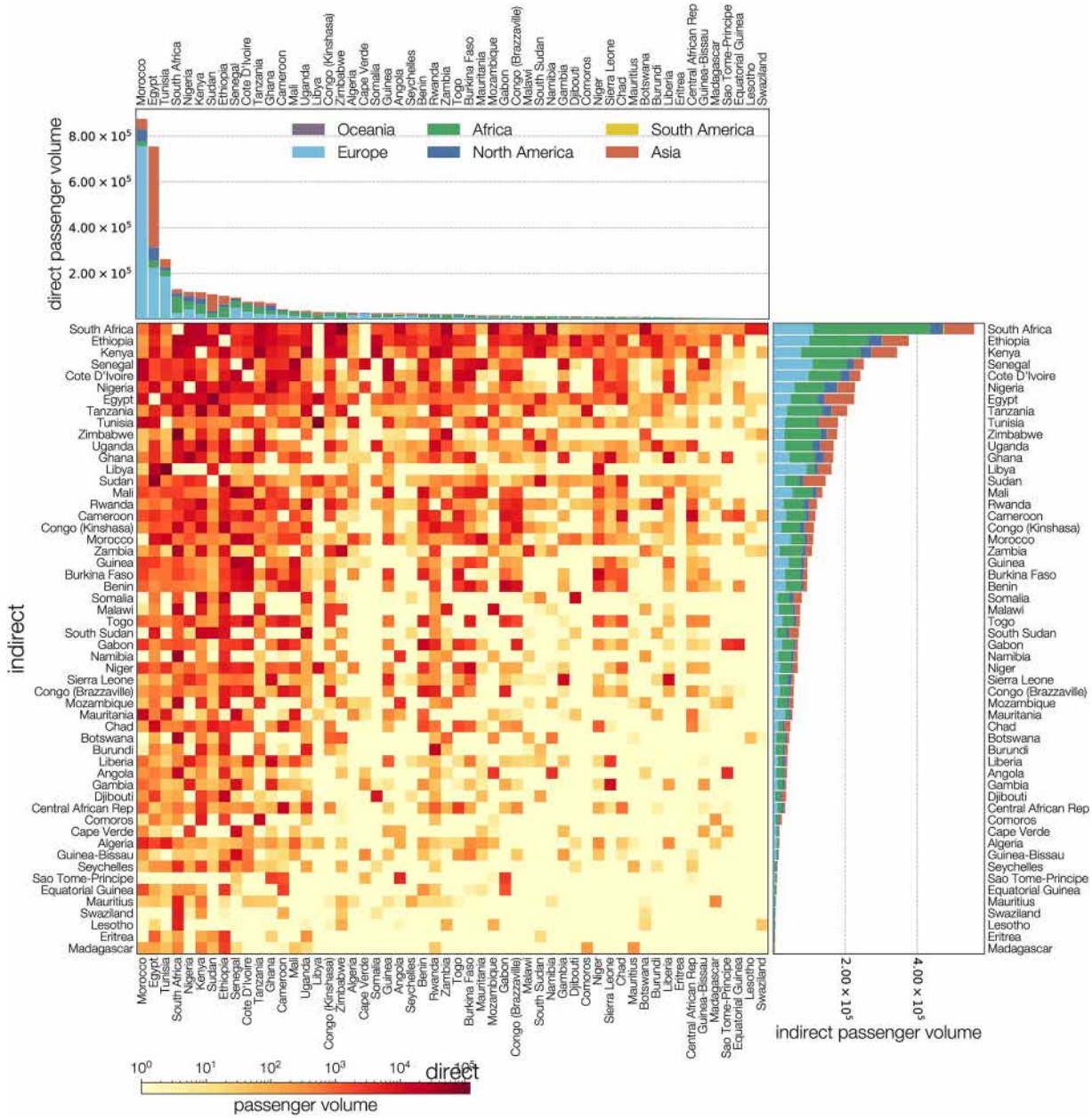

Figure 7. Patient data concerning symptoms, vaccination status and health outcomes. We show the symptoms reported for each Belgian patient infected with B.1.619, coloured according to the number of symptoms (of each patient; each row represents one patient). The three columns on the right show whether or not patients were hospitalised, and if they received oxygen therapy or spend time in the intensive care unit (ICU). The stars indicate patients that were vaccinated, with the two red colored stars indicating cases of post-vaccination infection.

new rules and regulations, for example requiring proof of vaccination or a negative test result upon departure and/or arrival by air. Increasingly, there have been criticisms concerning a lack of effectiveness of these measures, and concerning their economical ramifications. For example, with the emergence of the Omicron variant, some countries were quick to implement travel restrictions on various African countries. This included Ghana, given that the country had reported cases of the variant. However, the Ghanaian cases had been sampled from international travellers, prompting protest from the government about the travel ban. The economic ramifications of such a ban might dissuade countries from reporting new cases. Studies have found that, while border closure might have been effective in the beginning stages of the pandemic, this effectiveness waned once SARS-CoV-2 established local transmission chains or was limited (Grubaugh et al., 2019; Shi et al., 2020; Shiraef et al., 2022). Other studies showed that travel restrictions were simply not sufficient to prevent introductions from occurring (Hodcroft et al., 2021).

Note that for most pairs of countries in our study, the frequency of flights is relatively low, which is especially noticeable when contrasted to some other pairs for which the volume of flights is orders of magnitude higher. One example are flights between the United States and Canada. Meanwhile, Cameroon shows consistently low flight numbers throughout our study period, even though it plays an important role in the spread of B.1.619. This trend can also be seen in Figure 8, where we can see that some countries in Africa are responsible for a significant amount of international and intercontinental air travel compared

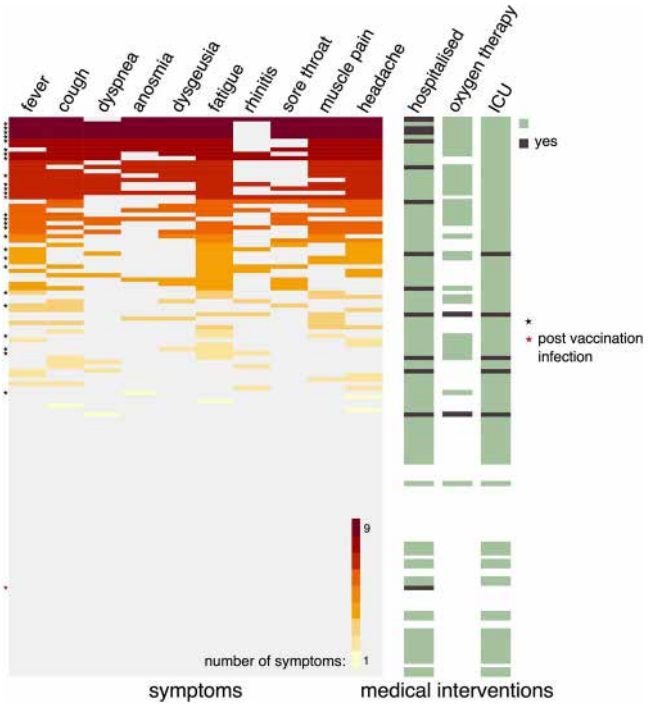

Figure 8. Within-Africa air passenger flux matrix and outward flux (from African countries) to the rest of the world. Columns of the matrix represent destination locations within Africa while the plot above the matrix indicates the continental breakdown of flights leaving that African country directly, *e.g.*. the vast majority of passengers flying from Morocco travel directly to Europe. Rows of the matrix represent origin locations within Africa. The plot to the right of the matrix indicates, for every row country, which continents it is likely to connect with secondarily, *e.g.* flights departing South Africa mostly land in countries that later send most of their passengers to African countries, while for Sudan it is Asia and for Libya it is Europe.

to others. Morocco sees a lot of flights going towards Europe, while Egypt mostly has outbound flights towards Asia. It is good to keep in mind that flight data is also an economic metric, and that the number of flights will often be higher for more economically developed nations. In the case of B.1.619, the likely origin is Africa, a region historically less covered by international air travel.

Genetic surveillance of B.1.619

The role of genetic surveillance in the context of the SARS-CoV-2 pandemic can not be understated, as it allows for timely detection of concerning lineages. This is illustrated by the discovery and subsequent global spread of the B.1.1.529 strain (also denominated Omicron). From a phylogenetic point of view, the Omicron variant is characterised by its genetic distance from other strains circulating at that time. Phylogenetic analyses indeed show the Omicron clade exhibiting a very long branch length, and it contains a high number of mutations compared to other variants (with over 30 in its spike protein alone) (Sun et al., 2022; Viana et al., 2022). This genetic distance raises questions about how to further improve genetic surveillance strategies, how is it possible that no earlier Omicron sequences were detected? Some researchers hypothesise that it circulated undetected in an animal reservoir, or that it developed in an immunocompromised person. The possibility that the development of Omicron simply escaped our notice should also not be fully excluded.

International cooperation in genetic surveillance

International efforts should be made in order to improve genetic surveillance systems in developing nations (Brito et al., 2022). However, political factors can play an important role in determining the efficiency of such surveillance efforts. For example, in Tanzania, former president John Pombe Magufuli had expressed scepticism about the existence of SARS-CoV-2 and declared the country COVID-free in mid-2020 (Buguzi, 2021). As of this writing, only 11 SARS-CoV-2 sequences from Tanzania have been uploaded on GISAID, even though media articles have reported outbreaks (Rice et al., 2021; Tarimo and Wu, 2020). Missing or incomplete data constitute an issue when performing phylogenetic and phylogeographic analyses, especially in cases where time (to results) is crucial, such as during an ongoing pandemic. In the case of B.1.619, new sequences were still being uploaded to GISAID after our analyses had been completed, many of them being Canadian, Turkish or Cameroonian in origin. In the case of many of these (more) recently uploaded sequences, the collection dates suggest that these samples are being processed a few months after being collected. The problems associated with data availability are well known, but are difficult to solve in a global effort.

While international cooperation and surveillance are of the utmost importance, analyses within one country can also be illuminating. In our analysis of Belgian sequences, we observed a few clusters, but a lot of relatively long distance jumps within the country. Considering the size of Belgium, the relatively long geographical distance of these jumps is not very surprising. Observing these patterns of transmission within a country for a single lineage or mutation can inform public health officials and allow them to act quickly, especially if such analyses can be conducted rapidly (Bollen et al., 2021; Dellicour et al., 2020).

The challenge of record keeping

A challenge that is often posed with regard to genomic surveillance is the completeness and correctness of sequence metadata. Often, the task of gathering and noting these data falls on medical personnel in the field, who are not always briefed to understand the importance of such metadata, and whose resources may already be fully utilized by the care of their patients. When trying to summarise the metadata for the Belgian sequences, we encountered this problem. Often, there was no distinction between a cell having been left blank because of a lack of information (i.e. missing data), or it being left blank to signify the lack of a symptom. If it is indicated that a patient had a fever and fatigue, and nothing else is indicated, then the most likely assumption is that whoever noted these symptoms meant to convey that fever and fatigue were the only symptoms that the patient had. However, the problem is that this remains an assumption. Educating on the importance of clear and precise record-keeping could help prevent such difficulties. Another measure that could be effective is the implementation of standardised and easy-to-use tools (for example a record-keeping software with a user-friendly interface) which would further improve the quality of data and reduce the amount of work that needs to be done by medical personnel. Furthermore, extra personnel, less strict regulation and better pay could also incentivise and aid healthcare workers in these efforts.

Mutations of concern

As the immunity of the population increases, so does the selective pressure on the virus to become more contagious and infectious.

As such, we expect to see more such mutations in the future. In the case of B.1.619, the chances of this strain establishing a large clade seems unlikely, given that this did not happen until now and the lineage is most likely extinct. However, given the genetic characteristics of B.1.619, it does not seem a luxury to maintain a certain level of surveillance for new lineages with similar mutations. Especially in areas with lower genetic surveillance coverage, there is still a need to gather and process genetic data more quickly.

Materials and methods

Study design

This study was initiated upon detection of SARS-CoV-2 strains in Switzerland in mid January 2021, bearing spike protein amino acid substitutions E484K and N440K in the receptor-binding domain (RBD), as well as mutations I210T, A879S, D936N, S939F and T1027I. SARS-CoV-2 genomes carrying these mutations form a well-defined cluster, leading them to be classified as lineage B.1.619 by Pangolin (Rambaut et al., 2020; O'Toole et al., 2021), with this novel classification being integrated into GISAID (Shu and McCauley, 2017) shortly after. We downloaded all available sequences of this lineage from GISAID on May 10, 2021, and subsequently reassessed which genomes belonged to this lineage. Briefly, due to occasional incorrect classification of sequences by Pangolin, sequences designated as B.1.619 on GISAID were combined with reference sequence NC_045512 and aligned using MAFFT (Katoh and Standley, 2013b) using the FFT-NS-2 setting. Insertions relative to the reference sequence were stripped and B.1.619 genomes were identified via amino acid changing mutations 6633 (ORF1a:A2123V), 8084 (ORF1a:E2607K), 11521 (ORF1a:M3752I), 22882 (S:N440K), 23012 (S:E484K), 24368 (S:D936N), 24378 (S:S939F), 24642 (S:T1027I), 26767 (M:I82T), 28875 (N:S201I), 28887 (N:T205I), and deletion 11,288-11,296 (ORF1a:SGF3675/3677). Genomes that did not contain these mutations were removed from the data set. We removed one Belgian genome corresponding to a sample being sequenced twice (EPI_ISL_1620233 and EPI_ISL_1620229). A GISAID acknowledgment table containing all genome accession numbers is included in Supplementary Materials.

Associated travel history

As in previous studies [e.g. (Lemey et al., 2020; Dudas et al., 2021)], we incorporated travel history into the Bayesian phylogeographic analyses. Not all countries have equal access to wide-scale genomic sequencing. In such cases, a lack of sampling can lead to an outbreak going undetected until it escapes its location of origin, with little to no data being available about the initial outbreak. Inclusion of individual travel histories allows for more realistic hypotheses of virus spread and a higher posterior predictive accuracy compared to including only sampling location (as in standard discrete phylogeography) (Lemey et al., 2020).

For B.1.619, we were able to obtain individual travel histories for seven infected travellers: 5 traveling out of Cameroon, 1 out of Rwanda and 1 out of South Korea. We obtained this information in part from the available metadata listed in GISAID, but mostly through an international network of collaborators. Since B.1.619 was first sampled in Switzerland, in travellers returning form Cameroon, Cameroon was suspected to be the location of origin for this lineage. Note that a closely related lineage (B.1.620) was also traced back to Cameroon (Dudas et al., 2021). As such, it was especially interesting to take into account travel history for this particular lineage. Table 1 shows the available travel information used in this study.

Due to privacy reasons, we were not able to obtain information concerning the precise dates of travel for the patients associated with these samples. Instead, as in Lemey et al. (2020) and Dudas et al. (2021), we treated the time of travel as a random variable and inferred the time of travel from the data (see the next section for the list of prior distributions used on all inferred parameters).

Phylogenetic and phylogeographic analysis

First, given the presumed origin of the B.1.619 lineage, we combined the available B.1.619 genomes with all genomes from an Africa-focused Nextstrain build available on June 25th 2021 (Hadfield et al., 2018). On this data set, we performed a maximum-likelihood (ML) phylogenetic analysis using IQ-TREE v2 (Minh et al., 2020) under a GTR+F+R5 model, as determined by IQ-TREE's automated model selection procedure. In the resulting phylogeny, all B.1.619 sequences, with the exception of one, were contained in one monophyletic clade. The one outlier sequence was an Ecuadorean sequence which instead closely clustered with other Ecuadorean genomes that were part of the collection of background genomes (from the Africa-focused Nextstrain build). This sequence has since been re-classified as belonging to the B.1 lineage. We selected a large subtree which contained the B.1.619 clade (see Figure 3), resulting in a data set of 1266 sequences in total, including 548 sequences not belonging to B.1.619. Including these non-B.1.619 sequences serves to provide context for the outbreak of B.1.619 within the larger global outbreak of SARS-CoV-2, and hence to aid with properly calibrating the phylogeny in time. To this end, we also decided to include an early sequence from Wuhan, China, (GISAID ID: EPI_ISL_406798) from December 2019. Finally, we re-aligned the full data set of 1266 sequences using MAFFT (Katoh and Standley, 2013a).

Second, we used this full data set to perform a discrete Bayesian phylogeographic analysis on the continent level using BEAST v1.10.5 (Suchard et al., 2018) with BEAGLE v3.2 (Ayres et al., 2019). We provided a starting tree for this analysis by first estimating a time-calibrated ML phylogeny using IQ-TREE v2 (using the same GTR+F+R5 model) and TreeTime v0.8.1 (without removing any outliers). In this step, we estimated the geographic spread of SARS-CoV-2 between continents: Asia ($n = 163$), Africa ($n = 340$), Europe ($n = 610$), South America ($n = 15$), North America ($n = 128$) and Oceania ($n = 10$), taking into account travel history data between these continents using a recently developed methodology (Lemey et al., 2020).

In a third step, we focused exclusively on the B.1.619 sequences within our data set. We again performed a discrete phylogeographic analysis using BEAST v1.10.5 (Suchard et al., 2018) with BEAGLE v3.2 (Ayres et al., 2019), but this time on the country level, using only the clade containing all 717 B.1.619 sequences (see Figure 3). In total, our data set contained B.1.619 genomes for 22 locations within this clade: Belgium ($n = 107$), Cameroon ($n = 20$), Canada ($n = 71$), Croatia ($n = 1$), Ecuador ($n = 1$), France ($n = 111$), Gabon ($n = 1$), Germany ($n = 161$), Ireland ($n = 2$), Japan ($n = 1$), Luxembourg ($n = 30$), Poland ($n = 2$), Russia ($n = 1$), Rwanda ($n = 1$), Slovakia ($n = 1$), South Korea ($n = 87$), Sweden ($n = 1$), Switzerland ($n = 74$), Turkey ($n = 23$), United Kingdom ($n = 5$), and USA ($n = 17$). In order to restrain the number of locations in the analysis (for computational reasons and to avoid statistical mixing issues), we removed locations with less than five sequences, with three exceptions. First, we included African countries with less

than five sequences (Gabon and Rwanda), since Africa is the suspected origin of this strain, and we wish to avoid removing any information surrounding the hypothesised start of the epidemic. Another exception was made for the Japanese sequence, since this sequence had an associated travel history out of South Korea, we found it of interest to include it. Finally, one Belgian sequence had an associated travel history out of South Africa, so this country was also added as a possible location to the discrete phylogeographic analysis. This resulted in a total of 14 countries being considered in our analysis, for a total data set size of 704 sequences.

For this country-level phylogeographic analysis, we parameterised the migration rates between countries as a log-linear function of a set of predictors, using a generalised linear model (GLM; Lemey et al. (2014)). We selected three predictor variables: (i) an air passenger flux matrix between all pairs of countries, containing total air passenger flow between these countries, from the start of October 2020 until the end of June 2021, obtained from the International Air Transport Association (IATA)(intelligence platform, xxxx); (ii) a distance matrix (containing the great circle distance between the centroids for each pair of countries); and (iii) a matrix that indicates whether a pair of countries share a border or not (Dudas et al., 2017). For the air passenger flux matrix, we considered both direct and indirect flights between pairs of countries. Given the implementation of travel restrictions by many countries at various points during the pandemic, we expected stark month-to-month variation in the number of flights between some pairs of countries. We can see this in Figure 8 and Supplementary Figure 11, where in some months the number of flights between certain pairs of countries decreases drastically. For example, the number of flights from France to the USA roughly halved in November 2021 compared to the month before. For this reason, we also included an analysis using an epoch model (Bielejec et al., 2014) where a different GLM is used for every month between October 2020 and June 2021, taking into account the air passenger flux between countries on a monthly basis during that specific time interval. For each of these phylogeographic analyses, we estimated the expected number of transitions between continents [known as Markov jumps; Minin and Suchard (2008a); Minin and Suchard (2008b)].

As the sampling date was unknown for a number of sequences, we included a prior on the tip dates based on their uncertainty, for all analyses performed. If only the year of sampling was known, we assumed a uniform prior between the first of January of that year and the date the sequence was uploaded. If only the month was known, we assumed a uniform prior between the first and last day of the month. The prior distributions in the analyses were the following: a gamma (shape = 0.001; scale = 1000) prior on the skygrid precision parameter, Dirichlet(1.0, K) priors on all sets of frequencies (with K the number of categories), gamma prior distributions (shape = rate = 1.0) on the unnormalised transition rates between locations (Lemey et al., 2009), a Poisson prior (country level: $\lambda = 14$; continent level: $\lambda = 5$) on the sum of non-zero transition rates between locations, and a CTMC reference prior on the mean evolutionary rate and as well as on the overall (constant) diffusion rate (Ferreira and Suchard, 2008). For the country-level analysis we used the estimated node height for the B.1.619 clade obtained during the continent-level analysis in order to define a normal prior on the root height (mean = 0.6, standard deviation = 0.13). For the travel times of infected passengers, we assumed a uniform interval with a mean length of 10 days before sampling time, with a standard deviation of 3 days. This is based on an estimated incubation time of 5 days, and another estimated period of 5 days between symptom onset and testing (Lauer et al., 2020a; Lauer et al., 2020b).

Table 1. Individual travel histories collected for the core genomic data set analysed in this study, sorted by the samples' collection dates. Importantly, most of the documented travel cases from Cameroon to several European countries were retrieved by contacting the labs that submitted the genomes to GISAID.

GISAID accession ID	Sampling location	Sampling date	Travel location
934981	Switzerland	2021-01-16	Cameroon
1302853	Switzerland	2021-02-09	Cameroon
1212838	Germany	2021-02-15	Cameroon
2361092	South Korea	2021-02-18	Cameroon
1312169	Belgium	2021-02-21	South Africa
1282092	Germany	2021-03-02	Cameroon
1568404	Germany	2021-03-30	Rwanda
2131789	Japan	2021-05-01	South Korea
2179659	Belgium	2021-05-15	France

Finally, we performed a continuous phylogeographic analysis using solely Belgian sequences. Following a previously developed workflow (Dellicour et al., 2020; Bollen et al., 2021), we first parsed the maximum clade credibility (MCC) tree obtained in the previous step as a way to identify introduction events into Belgium. We selected any clade within this larger MCC tree consisting of at least three genomic sequences with Belgian origin. For every Belgian sequence in our sample ($n = 107$), we knew the location on a municipality level and we assigned a random point within the municipality as the geographic coordinates for each sequence (Dellicour et al., 2020). Given the average size of Belgian municipalities is $53\mathbf{km}^2$, this still provided fine-grained data. For every clade of at least three Belgian sequences, we then performed a spatially-explicit phylogeographic analysis using a relaxed random walk (RRW) diffusion model in BEAST v1.10.5 (Lemey et al., 2010). We used a Cauchy distribution to model the among-branch heterogeneity in diffusion velocity. We collected a sample of 1000 posterior trees as input for the R package 'seraphim' (Dellicour et al., 2016) to extract the spatiotemporal information from these trees.

Acknowledgements

We gratefully acknowledge the authors from originating laboratories responsible for obtaining the specimens, as well as submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.
An acknowledgment table with accessions of SARS-CoV-2 genomes used here is included. We thank all involved in the collection and processing of SARS-CoV-2 testing and genomic data, as well as associated metadata on individual travel histories. We would also like to thank all sequencing laboratories for uploading SARS-CoV-2 genomes to the international GISAID database. In particular, we would like to thank the members of the COVID-19 Genomics Belgium consortium: Emmanuel André, Piet Maes, Guy Baele, Simon Dellicour, Lize Cuypers, Marc Van Ranst, Barney Potter, Samuel Hong, François E. Dufrasne, Guillaume Bayon-Vicente, Ruddy Wattiez, Carl Vael, Lynsey Berckmans, Philippe Selhorst, Kevin K. Ariën, Arnaud Marchant, Coralie Henin, Benoit Haerlingen, Ricardo De Mendonca, Marie-Luce Delforge, Sonia Van Dooren, Bruno Hinckel, Hideo Imamura,

Toon Janssen, Ben Caljon, Oriane Soetens, Denis Piérard, Thomas Demuyser, Charlotte Michel, Olivier Vandenberg, Sigi van den Wijngaert, Giulia Zorzi, Jean Ruelle, Benoit Kabamba Mukadi, Jean-Luc Gala, Bertrand Bearzatto, Jérôme Ambroise, Philippe Van Lint, Walter Verstrepen, Reinout Naesens, Michael Peeters, Kate Bakelants, Sarah Denayer, Sofieke Klamer, Pascale Hilbert, Sylvain Brohée, Pierre-Emmanuel Léonard, Deniz Karadurmus, Jeremie Gras, Damien Féret, Barbara Lambert, Anne Vankeerberghen, Astrid Holderbeke, Hans De Beenhouwer, Lien Cattoir, Christine Lammens, Basil Britto Xavier, Marie Le Mercier, Jasmine Coppens, Veerle Matheeussen, Herman Goossens, Geert A. Martens, Koen Swaert, Frederik Van Hoecke, Dieter Desmet, Pierre Bogaerts, Jonathan Degosserie, Olivier Denis, TeDin Huang, Dagmar Obbels, Hanne Valgaeren, Johan Frans, Annick Smismans, Paul-Emile Claus, Truus Goegebuer, Ann Lemmens, Bea Van den Poel, Sonja De Bock, Wim Laffut, Ellen Van Even, Jos Van Acker, Charlotte Verfaillie, Elke Vanlaere, Klara De Rauw, Brigitte Maes, Guy Froyen, Bert Cruys, Ellen Geerdens, Luc Waumans, Britt Van Meensel, Reinoud Cartuyvels, Severine Berden, Marijke Raymaekers, Bruno Verhasselt, Cécile Meex, Keith Durkin, Laurent Gillet, Maria Artesi, Marie-Pierre Hayette, Sébastien Bontems, Vincent Bours, Claire Gourzonès, Olivier Ek, Fabrice Bureau, Jorn Hellemans, Patrick Descheemaeker, and Marijke Reynders.

SLH acknowledges support from the Research Foundation—Flanders ('Fonds voor Wetenschappelijk Onderzoek—Vlaanderen', G0D5117N). GB, SD and SLH acknowledge support from the Research Foundation—Flanders ('Fonds voor Wetenschappelijk Onderzoek—Vlaanderen', G0E1420N). GB and BP acknowledge support from the Internal Funds KU Leuven (Grant No. C14/18/094). GB, SD and NB acknowledge support from the Research Foundation—Flanders ('Fonds voor Wetenschappelijk Onderzoek—Vlaanderen', G098321N). SD also acknowledges support from the 'Fonds National de la Recherche Scientifique' (F.R.S.-FNRS, Belgium; grant n°F.4515.22) and from the European Union Horizon 2020 project MOOD (grant agreement n°874850). This work was supported by Agence Française de Développement through the AFROSCREEN project (grant agreement CZZ3209), coordinated by ANRS | Maladies infectieuses émergentes in partnership with Institut Pasteur and IRD. We would additionally like to thank members from the AFROSCREEN Consortium (https://www.afroscreen.org/en/network/) for their work and support on genomic surveillance in Africa.

Supplementary data

Supplementary data is available at VEVOLU Journal online.

Conflict of interest: NS, EG, RD, DB, MA, MB, NVI and MCP are employees of Philip Morris International. PD was contracted and paid by Philip Morris International.

Data availability statement

BEAST XML input files are available as supplementary files. The SARS-CoV-2 genome data required to run these XML files can be downloaded from https://www.gisaid.org; all GISAID accession numbers are part of the sequence names/identifiers.

References

Adebisi YA, Oke GI., Ademola PS *et al.* SARS-CoV-2 diagnostic testing in Africa: needs and challenges. *Pan African Med J* 2020;**35**:4, https://doi.org/10.11604/pamj.2020.35.4.22703.

Alteri C, Cento V, Piralla A *et al.* Genomic epidemiology of SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy. *Nat Commun* 2021;**12**:434. https://doi.org/10.1038/s41467-020-20688-x

Augusto G, Mohsen MO, Zinkhan S *et al.* In vitro data suggest that Indian delta variant b.1.617 of SARS-CoV-2 escapes neutralization by both receptor affinity and immune evasion. *Allergy* 2021;**77**:111–117.

Ayres DL, Cummings MP, Baele G *et al.* BEAGLE 3: Improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Syst Biol* 2019;**68**:1052–1061.

Bielejec F, Lemey P, Baele G *et al.* Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. *Syst Biol* 2014;**63**:493–504.

Boehm E, Kronig I, Neher RA *et al.* Novel SARS-CoV-2 variants: the pandemics within the pandemic. *Clin Microbiol Infect* 2021;**27**:1109–1117, https://doi.org/10.1016/j.cmi.2021.05.022.

Bollen N, Artesi M, Durkin K *et al.* Exploiting genomic surveillance to map the spatio-temporal dispersal of SARS-CoV-2 spike mutations in Belgium across 2020. *Sci Rep* 2021;**11**:18580. https://doi.org/10.1038/s41598-021-97667-9

Brito AF, Semenova E, Dudas G *et al.* Global disparities in SARS-CoV-2 genomic surveillance. *Nat Commun* 2022;**13**:5992. https://doi.org/10.1038/s41467-022-33713-y

Buguzi S Covid-19: Counting the cost of denial in Tanzania. *BMJ* 2021;**373**:n1052. https://doi.org/10.1136/bmj.n1052

Cuypers L, Dellicour S, Hong SL *et al.* Two years of genomic surveillance in Belgium during the SARS-CoV-2 pandemic to attain country-wide coverage and monitor the introduction and spread of emerging variants. *Viruses* 2022;**14**:2301. https://doi.org/10.3390/v14102301

N. G. Davies, S. Abbott, R. C. Barnard, *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage b.1.1.7 in England. *Science*, **372**(6538):eabg3055, 2021. https://doi.org/10.1126/science.abg3055.

Dellicour S, Rose R, Faria NR *et al.* SERAPHIM: studying environmental rasters and phylogenetically informed movements. *Bioinformatics* 2016;**32**:3204–3206. https://doi.org/10.1093/bioinformatics/btw384

Dellicour S, Durkin K, Hong SL *et al.* A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *Mol Biol Evol* 2020;**38**:1608–1613, https://doi.org/10.1093/molbev/msaa284.

Dudas G, Carvalho LM, Bedford T *et al.* Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 2017;**544**:309–315.

Dudas G, Hong SL, Potter BI. *et al.* Emergence and spread of SARS-CoV-2 lineage B.1.620 with variant of concern-like mutations and deletions. *Nat Commun* 2021;**12**:5769. https://doi.org/10.1038/s41467-021-26055-8

Ferreira MAR, Suchard MA Bayesian anaylsis of elasped times in continuous-time Markov chains. *Can J Stat* 2008;**26**:355–368.

Frampton D, Rampling T, Cross A *et al.* Genomic characteristics and clinical effect of the emergent SARS-CoV-2 b.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect Dis* 2021;**21**:1246–1256, https://doi.org/10.1016/s1473-3099(21)00170-5.

Funk T, Pharris A, Spiteri G *et al.* Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance* 2021;**26(16)**:2100348. https://doi.org/10.2807/1560-7917.ES.2021.26.16.2100348

Greaney AJ, Loes AN, Crawford KHD *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 2021;**29**:463–476.e6, https://www.cell.com/cell-host-microbe/abstract/S1931-3128(21)00082-2.

Grubaugh ND, Saraf S, Gangavarapu K *et al.* Travel surveillance and genomics uncover a hidden zika outbreak during the waning epidemic. *Cell* 2019;**178**:1057–1071.e11, https://doi.org/10.1016/j.cell.2019.07.018.

Hadfield J, Megill C, Bell SM *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;**34**:4121–4123, https://doi.org/10.1093/bioinformatics/bty407.

Hamer DH, Barbre KA, Chen LH *et al.* Travel-associated zika virus disease acquired in the Americas through February 2016. *Ann Intern Med* 2016;**166**:99, https://doi.org/10.7326/m16-1842

Harvey K, Esposito DH, Han P *et al.* Surveillance for travel-related disease–GeoSentinel Surveillance System, United States, 1997-2011. *MMWR Surveill Summ* 2013;**62**:1–23.

Harvey WT, Carabelli AM, Jackson B *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;**19**:409–424.

Hodcroft EB, Zuber M, Nadeau S *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* 2021;**595**:707–712, https://doi.org/10.1038/s41586-021-03677-y

Hoffmann M, Krüger N, Schulz S *et al.* The omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic. *Cell* 2022;**185**:447–456.

B. O. intelligence platform. https://bluedot.global/.

Jangra S, Ye C, Rathnasinghe R *et al.* SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *Lancet Microbe* 2021;**2**:e283–e284. https://doi.org/10.1016/s2666-5247(21)00068-9

Jaspe RC, Sulbaran Y, Loureiro CL *et al.* Importance of E484K and N501Y mutations in SARS-CoV-2 for genomic surveillance: rapid detection by restriction enzyme analysis. *medRxiv* 2021;Preprint. https://doi.org/10.1101/2021.05.04.21256650

Jung C, Kmiec D, Koepke L *et al.* Omicron: what makes the latest SARS-CoV-2 variant of concern so concerning?. *J Virol* 2022;**96**:e02077–21.

Kaleta T, Kern L, Hong SL *et al.* Antibody escape and global spread of SARS-CoV-2 lineage A.27. *Nat Commun* 2022;**13**:1–13.

Katoh K, Standley DM MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013a;**30**:772–780, https://doi.org/10.1093%2Fmolbev%2Fmst010.

Katoh K, Standley DM Mafft multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013b;**30**:772–780.

Lau H, Khosrawipour V, Kocbach P *et al.* The association between international and domestic air traffic and the coronavirus (covid-19) outbreak. *J Microbiol Immunol Infect* 2020;**53**:467–472, https://doi.org/10.1016/j.jmii.2020.03.026.

S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Technical report*, Imperial College London, 2020a.

Lauer SA, Grantz KH, Bi Q *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020b;**172**:577–582.

Leder K, Torresi J, Brownstein JS *et al.* Travel-associated illness trends and clusters, 2000–2010. *Emerg Infect Dis* 2013;**19**:1049–1073, https://doi.org/10.3201/eid1907.121573.

Lemey P, Rambaut A, Drummond AJ, *et al.* Bayesian phylogeography finds its roots. *PLoS Comp Biol* 2009;**5**:e1000520.

Lemey P, Rambaut A, Welch JJ, *et al.* Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 2010;**27**:1877–1885, https://doi.org/10.1093/molbev/msq067.

Lemey P, Rambaut A, Bedford T *et al.* Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Path.* 2014;**10**:e1003932.

Lemey P, Hong SL, Hill V *et al.* Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat Commun* 2020;**11**:5110. https://doi.org/10.1038/s41467-020-18877-9

McCallum M, Marco AD, Lempp FA *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 2021;**184**:2332–2347.e16, https://doi.org/10.1016/j.cell.2021.03.028

Minh B, Schmidt H, Chernomor O *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–1534.

Minin V, Suchard MA Counting labeled transitions in continuous-time Markov models of evolution. *J Math Biol* 2008a;**56**:391–412.

Minin V, Suchard MA Fast, accurate and simulation-free stochastic mapping. *Phil Trans R Soc B Biol Sci* 2008b;**363**:3985–3995.

Nonaka CKV, Franco MM, Gräf T *et al.* Genomic evidence of a sars-cov-2 reinfection case with e484k spike mutation in brazil. *Preprints* 2021 https://doi.org/10.20944/preprints202101.0132.v1

O'Toole A, Scher E, Underwood A *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* 2021;**7(2)**:veab064. https://doi.org/10.1093/ve/veab064

Park AK, Kim I-H, Kim HM *et al.* SARS-CoV-2 b.1.619 and b.1.620 lineages, south korea, 2021. *Emerg Infect Dis* 2022;**28**:415–419, https://doi.org/10.3201/eid2802.211653.

Planas D, Veyer D, Baidaliuk A *et al.* Reduced sensitivity of SARS-CoV-2 variant delta to antibody neutralization. *Nature* 2021;**596**:276–280. https://doi.org/10.1038/s41586-021-03777-9

Rambaut A, Holmes EC, O'Toole A *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;**5**:1403–1407.

Rice BL, Annapragada A, Baker RE *et al.* Variation in SARS-CoV-2 outbreaks across sub-saharan africa. *Nat Med* 2021;**27**:447–453, https://doi.org/10.1038/s41591-021-01234-8.

Ruis C Sub-lineage of B.1.619 predominantly in South Korea. https://github.com/cov-lineages/pango-designation/issues/160, 2022. (Accessed: 23 November 2022)

Salyer SJ, Maeda J, Sembuche S *et al.* The first and second waves of the covid-19 pandemic in africa: a cross-sectional study. *The Lancet* 2021;**397**:1265–1275, https://www.sciencedirect.com/science/article/pii/S0140673621006322

Shi S, Tanaka S, Ueno R *et al.* Travel restrictions and SARS-CoV-2 transmission: an effective distance approach to estimate impact. *Bull World Health Organization* 2020;**98**:518–529. https://doi.org/10.2471/blt.20.255679

Shiraef MA, Friesen P, Feddern L *et al.* Did border closures slow SARS-CoV-2?. *Sci Rep* 2022;**12**:1709. https://doi.org/10.1038/s41598-022-05482-7

Shu Y, McCauley J GISAID: Global initiative on sharing all influenza data - from vision to reality. *EuroSurveillance* 2017;**22(13)**:30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Sigal A Milder disease with omicron: is it the virus or the pre-existing immunity?. *Nat Rev Immunol* 2022;**22**:69–71, https://doi.org/10.1038/s41577-022-00678-4.

Sikora M, von Bülow S, Blanc FEC *et al.* Computational epitope map of sars-cov-2 spike protein. *PLOS Comput Biol* 2021;**17**:1–16, https://doi.org/10.1371/journal.pcbi.1008790

Sokadjo YM, Atchadé MN The influence of passenger air traffic on the spread of COVID-19 in the world. *Transp Res Interdiscip Perspect* 2020;**8**:100213. https://doi.org/10.1016/j.trip.2020.100213

Suchard MA, Lemey P, Baele G *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;**4**:vey016. https://doi.org/10.1093/ve/vey016

Sun Y, Lin W, Dong W *et al.* Origin and evolutionary analysis of the SARS-CoV-2 omicron variant. *J Biosafety Biosecurity* 2022;**4**:33–37. https://doi.org/10.1016/j.jobb.2021.12.001

Tandel D, Gupta D, Sah V *et al.* N440K variant of SARS-CoV-2 has higher infectious fitness. *bioRxiv* 2021;**2021**:441434. https://doi.org/10.1101/2021.04.30.441434

Tarimo CS, Wu J The first confirmed case of COVID-19 in Tanzania: recommendations based on lesson learned from China. *Trop Med Health* 2020;**48**:25. https://doi.org/10.1186/s41182-020-00214-x

Viana R, Moyo S, Amoako DG *et al.* Rapid epidemic expansion of the SARS-CoV-2 omicron variant in Southern Africa. *Nature* 2022;**603**:679–686, https://doi.org/10.1038/s41586-022-04411-y

WHO. Who coronavirus (covid-19) dashboard. https://covid19.who.int/, 2022. (Accessed: 1 December 2022).

Wilder-Smith A, Quam M, Sessions O *et al.* The 2012 dengue outbreak in Madeira: exploring the origins. *Eurosurveillance* 2014;**19**:20718. https://doi.org/10.2807/1560-7917.ES2014.19.8.20718

Yang W, Zhang W, Kargbo D *et al.* Transmission network of the 2014–2015 ebola epidemic in sierra leone. *J R Soc Interface* 2015;**12**:20150536, https://doi.org/10.1098/rsif.2015.0536.

Virus Evolution, 2025, **11**, 1–14
https://doi.org/10.1093/ve/veaf017
Research Article