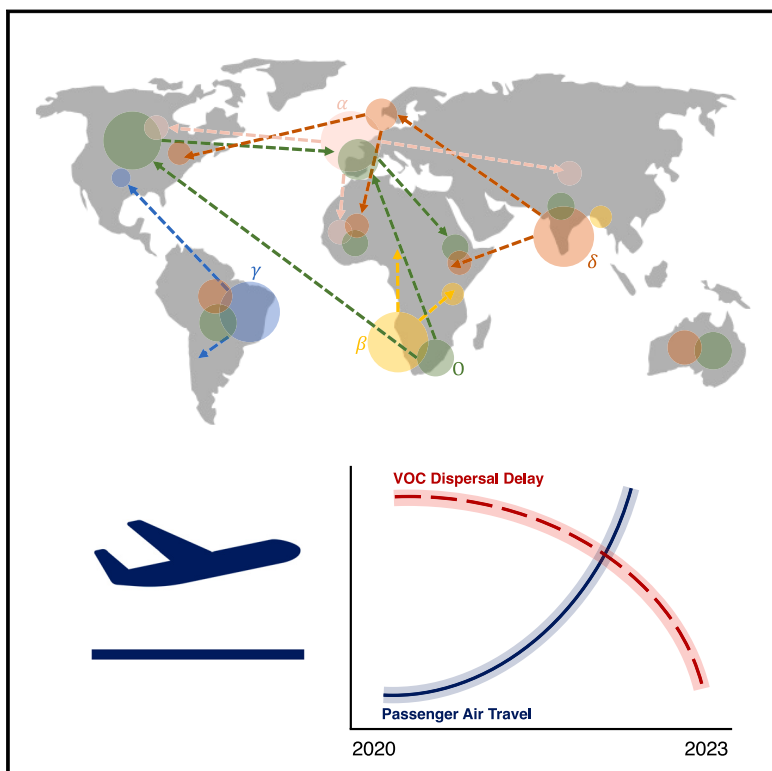# Dispersal patterns and influence of air travel during the global expansion of SARS-CoV-2 variants of concern

Graphical abstract

Authors

Houriiyah Tegally, Eduan Wilkinson, Joseph L.- H. Tsui, ..., Nuno R. Faria, Moritz U.G. Kraemer, Tulio de Oliveira

Correspondence

houriiyah.tegally@gmail.com (H.T.), moritz.kraemer@biology.ox.ac.uk (M.U.G.K.), tulio@sun.ac.za (T.d.O.)

In brief

Data analysis clarifies that dispersal of SARS-CoV-2 variants from their sites of initial detection was related to the amount of global air travel at the time of the variant's emergence and that travel volume through "hub" sites distinct from the site of emergence was a key driver of variant spread.

Highlights

- Global phylogenetic analysis reveals dispersal of VOCs along worldwide flight network

- Omicron spread to five times more countries within 100 days of emergence than other VOCs

- Delta and Omicron dispersed from secondary hubs during times of accelerating air travel

- Highly connected countries were major global and regional exporters of VOCs

CelPress

Article

Dispersal patterns and influence of air travel during the global expansion of SARS-CoV-2 variants of concern

Houriiyah Tegally,[1,2,19,21,*] Eduan Wilkinson,[1,19] Joseph L.- H. Tsui,[3,19] Monika Moir,[1,19] Darren Martin,[4,5] Anderson Fernandes Brito,[6] Marta Giovanetti,[7,8,9] Kamran Khan,[10,11] Carmen Huber,[10] Isaac I. Bogoch,[11] James Emmanuel San,[2] Jenicca Poongavanan,[1] Joicymara S. Xavier,[1,12,13] Darlan da S. Candido,[14] Filipe Romero,[14] Cheryl Baxter,[1] Oliver G. Pybus,[3,15,16] Richard J. Lessells,[2] Nuno R. Faria,[3,14,17] Moritz U.G. Kraemer,[3,15,20,*] and Tulio de Oliveira[1,2,18,20,*]

[1]Centre for Epidemic Response and Innovation (CERI), School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa
[2]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa
[3]Department of Biology, University of Oxford, Oxford, UK
[4]Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Cape Town, South Africa
[5]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa
[6]Instituto Todos pela Saúde, São Paulo, São Paulo, Brazil
[7]Laboratorio de Flavivirus, Fundacao Oswaldo Cruz, Rio de Janeiro, Brazil
[8]Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[9]Department of Science and Technology for Humans and the Environment, University of Campus Bio-Medico di Roma, Rome, Italy
[10]BlueDot, Toronto, ON, Canada
[11]Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, ON, Canada
[12]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[13]Institute of Agricultural Sciences, Universidade Federal dos Vales do Jequitinhonha e Mucuri, Unaí, Brazil
[14]MRC Centre for Global Infectious Disease Analysis and Department of Infectious Disease Epidemiology, Jameel Institute, School of Public Health, Imperial College London, London, UK
[15]Pandemic Sciences Institute, University of Oxford, Oxford, UK
[16]Department of Pathobiology and Population Sciences, Royal Veterinary College London, London, UK
[17]Departamento de Moléstias Infecciosas e Parasitárias e Instituto de Medicina Tropical da Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil
[18]Department of Global Health, University of Washington, Seattle, WA, USA
[19]These authors contributed equally
[20]Senior author
[21]Lead contact
*Correspondence: houriiyah.tegally@gmail.com (H.T.), moritz.kraemer@biology.ox.ac.uk (M.U.G.K.), tulio@sun.ac.za (T.d.O.)
https://doi.org/10.1016/j.cell.2023.06.001

SUMMARY

The Alpha, Beta, and Gamma SARS-CoV-2 variants of concern (VOCs) co-circulated globally during 2020 and 2021, fueling waves of infections. They were displaced by Delta during a third wave worldwide in 2021, which, in turn, was displaced by Omicron in late 2021. In this study, we use phylogenetic and phylogeographic methods to reconstruct the dispersal patterns of VOCs worldwide. We find that source-sink dynamics varied substantially by VOC and identify countries that acted as global and regional hubs of dissemination. We demonstrate the declining role of presumed origin countries of VOCs in their global dispersal, estimating that India contributed <15% of Delta exports and South Africa <1%–2% of Omicron dispersal. We estimate that >80 countries had received introductions of Omicron within 100 days of its emergence, associated with accelerated passenger air travel and higher transmissibility. Our study highlights the rapid dispersal of highly transmissible variants, with implications for genomic surveillance along the hierarchical airline network.

INTRODUCTION

Since the emergence of SARS-CoV-2 in late 2019, multiple waves of infection have spread across the world. Successive waves have been caused typically by new variants, each of which replaced previously dominant variants due to higher transmissibility and/or ability to evade immunity. At the end of 2020, the first three variants of concern (VOCs), Alpha,[1] Beta,[2] and

Gamma[3] emerged. Along with some less successful novel lineages (termed variants of interest or VOIs), these VOCs were characterized by a combination of increased intrinsic transmissibility, sometimes enhanced immune evasion capabilities, and increased pathogenicity.[4] Each of the VOCs was associated initially with increasing SARS-CoV-2 incidence in their presumed countries of origin: Alpha in the United Kingdom,[1,5] Beta in South Africa[2] and Gamma in Brazil.[3] In the second quarter of 2021, these three VOCs started being displaced worldwide by Delta, a fourth VOC with increased intrinsic transmissibility and pathogenicity compared with the initial three VOCs. Retrospective analysis revealed that, as with Alpha and Beta, Delta may have first arisen between September and October 2020,[6] but only spread globally after it caused a large outbreak in India in March 2021.[7] The most recent VOC, Omicron, was detected first during a rapid increase in cases in Botswana and South Africa in November 2021.[8]

At each stage of the pandemic, global transmission of SARS-CoV-2 has continued within a context of shifting public health responses, virus evolution, and dynamic changes in host immunity. The pandemic precipitated unprecedented changes in the intensity and nature of human mobility, both internationally through strict restrictions on global travel and nationally via government-implemented public health and social measures (PHSMs).[9] Although the intention behind initial restrictions on international travel was to limit the dispersal of the virus out of the first outbreak epicenters with a view to possible elimination, they were later used to attempt to limit or slow the dispersal of VOCs out of their perceived regions of origin and to reduce epidemic intensity. Following the global vaccination rollout, travel restrictions and other PHSM were gradually lifted in most parts of the world, bringing travel and mobility patterns back toward levels seen before the pandemic.[10] Regardless of the intensity and range of travel restrictions, multiple SARS-CoV-2 variants have disseminated and risen to prominence across large swathes of the world.

Understanding the global dispersal patterns of SARS-CoV-2 VOCs in the context of local and global human mobility is critical if we are to objectively evaluate the relative importance of targeted travel restrictions as pandemic prevention and/or control measures. Fortunately, increased investments in genomic surveillance and data sharing throughout the pandemic have enabled the effective tracking of VOCs in near real time, mainly through the GISAID database (https://gisaid.org/).[11] Consequently, sufficient genomic sequence data are now available to enable detailed investigations of past SARS-CoV-2 transmission dynamics in different locations and at different scales.[6,12–17] Yet, the factors underlying variability in the dissemination of SARS-CoV-2 VOCs are yet to be fully understood, especially on a global scale and when comparing all five VOCs.

Here, we combine phylogenetic models that leverage multiple sets of ~20,000 genomes per VOC from >100 countries with global air passenger data in order to reconstruct the global spread of each VOC. We investigate the movement dynamics of each VOC at country and regional levels to determine source-sink dynamics and establish the regional and global contributions of individual countries to the exportation of VOCs. We specifically investigate the role of the countries that first reported

an epidemic of each VOC on the global movement dynamics of VOCs and measure the influence of international travel on VOC dispersal.

RESULTS

VOC global dissemination patterns

To quantify the global dissemination patterns of each VOC, we performed ancestral state reconstruction of discrete spatial locations using dated phylogenetic trees that were inferred from a subset of representative sampled genomes (for which sequence sampling locations were known). Genomes were sampled in proportion to country- and variant-specific case counts, and analyses were repeated across 10 replicates of ~20,000 randomly sampled genomes per VOC (after accounting for VOC-specific case counts). Continental source-sink dynamics were determined by calculating the net difference between viral exportation and importation events for each country and continent (see method details). Given the long duration of dissemination of certain VOCs, this analysis was performed both to give an overall assessment for the whole period (Figures 1 and 2) and to provide insights into the temporal heterogeneity in the directionality of dispersal (Figures S1A and S1B). A limiting factor of this analysis is that countries with under-reported incidence and low sequencing proportions,[18] but high global connectivity would have been missed as important global or regional VOC disseminators (given the reliance of our methods on genomic data and underlying testing patterns).

Our analyses reveal distinct global expansion processes for each VOC. The Alpha, Beta, and Gamma variants co-circulated globally from November and December 2020 to June and July 2021 (Figure S2). As expected, Europe was a major source of the Alpha variant, with the UK contributing the highest estimated number of relative exports to the rest of the world (>2,000; ~50% of Alpha exports; Figure 1) throughout its dissemination period (Figure S1B). Global expansion can be described as a multistage process—first, at the end of 2020 and beginning of 2021, we estimated that Alpha would spread mostly within Europe (>3,000 exports between European regions) and from Europe to the Americas and Asia (>600 exports from Europe). Most introductions of Alpha to Africa were from Europe or North America (>60 exports), initially to West Africa and then to East Africa (Figure 1). Second, between February and May 2021, Alpha spread within the Americas and Asia, and we observed viral lineage exchange between East Africa and Asia. During this time, North America acted as an overall source of Alpha, along with Europe (Figure S1B). We estimated that Africa's and South America's contributions to global dissemination of Alpha were minimal, overall. It is important to note that due to subsampling and uneven sampling, viral importation numbers presented in this manuscript need to be interpreted as relative measures and will underestimate the actual number of importations. Additionally, country-specific differences in testing rates and sequencing efforts could also introduce potential biases in the estimated numbers of international exports, particularly when testing is low or sequencing intensity is much lower in proportion to recorded cases, where our method would underestimate the numbers of viral imports and exports.
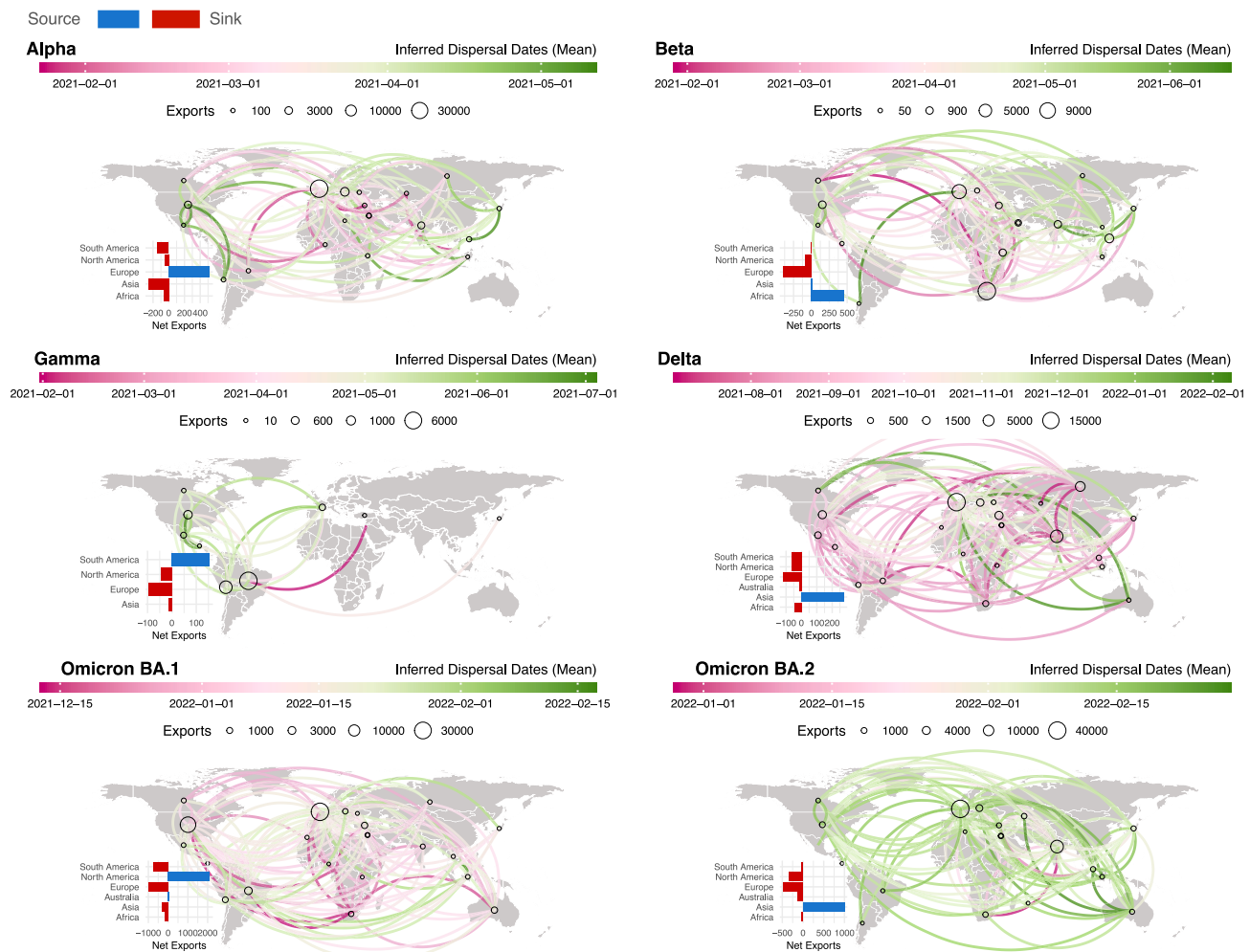
Figure 1. Spatiotemporal dispersal patterns of VOCs
Global dissemination and continental source-sink dynamics for each VOC, determined from ancestral state reconstruction analysis. Virus lineage exchanges are aggregated at the sub-continental level, and curves linking any two locations are colored according to the mean dates of all viral movements inferred along this route. Sub-continental level denominations vary by region, where in some regions they denote countries (e.g., the US and Canada in North America), whereas in others they denote groups of countries (e.g., western Europe). The curves denote the direction of movement in an anti-clockwise direction. Circles are drawn proportional to the number of exports per sub-continental location. Source and sink continents are determined by calculating the net difference between viral exportation and importation events. The absolute numbers of exportation and importation events for each continent per VOC are shown in Figures S3B and S3C.

Most Beta exportations were from southern Africa (∼1,200 estimated exports; ∼48% of Beta exports), and around half of those were to locations within the same region (>600 exports) (Figure 1). We also infer considerable Beta spread from southern Africa to western Europe (∼300 exports) and then within western Europe (∼400 exports). We infer that Asia was a net source of Beta along with Africa, with exports peaking after April 2021 for Asia (Figure S1B), which is plausible given sizable Beta waves in some countries in Asia during that time (e.g., Bangladesh and Cambodia).[19] We again observe a multi-stage process of Beta spread: during the earliest stages between late 2020 and early 2021, Beta was primarily exported from southern Africa into North America, Europe, and Asia. Later, dispersal from Asia occurred mostly to Europe and North America, with minimal introductions to South America (Figure 1).

Gamma circulation was first detected in Brazil, where it caused a large wave of infection between December 2020 and March 2021.[3] From there, Gamma was exported mainly to other South American countries (∼50% of Gamma exports) throughout the dissemination period (Figure S1B). Only later, between May and June 2021, do we infer a few instances of Gamma spread from North America to Europe or from Europe back to the Americas (Figure 1).

The global spread of Delta was characterized by early exportations from the Indian subcontinent and Russia to other regions of Asia and all other parts of the world during the first half of 2021 (Figure 1). Russia's estimated contribution to international Delta exports (∼11% of global Delta exports in our study) is consistent with previous work inferring dispersal from there based on human mobility data.[6] In the second half of 2021, many more

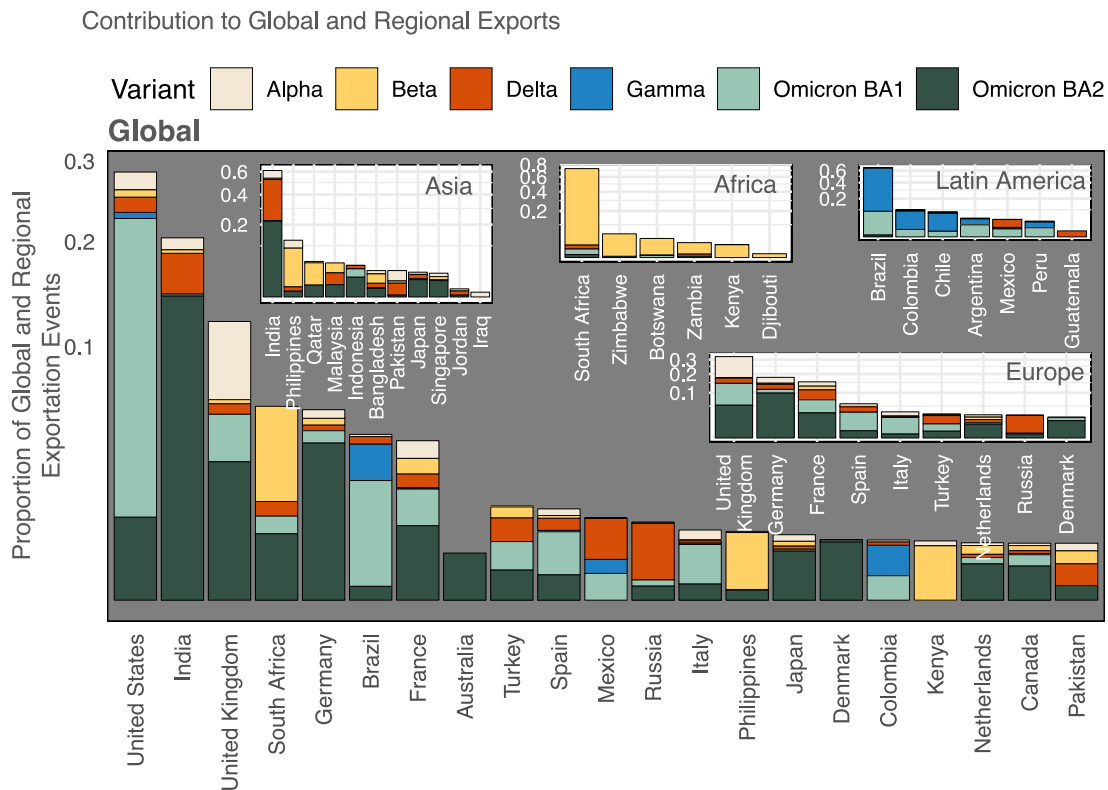## Contribution to Global and Regional Exports

Figure 2. Regional and global dissemination hubs of VOCs
Largest global and regional contributors to viral exports, stratified by VOC. Countries are shown here if they contribute >0.1% of exports globally or within Europe, or >0.5% within other regions.

inferred dispersal events originated from western Europe, including the UK (>1,100 within western Europe, >300 to central Europe, and ~30 to the USA, Brazil, and the Middle East, respectively; Figures 1 and S1B, also in line with previous work[6]). Although western Europe demonstrated the largest absolute number of exportation events of the Delta variant, Europe still acted as an overall sink (when balancing out comparably large importation numbers into Europe), and Asia remained a major net source when considering the net volumes of inbound, as well as outbound movements of Delta lineages (Figure 1). Given Delta's extended period of dissemination globally compared with Alpha, Beta, and Gamma, it is important to consider the time variability in exportation patterns for this VOC. From the results, we observe that, although the majority of exports were occurring from and within Asia in the initial months of the dispersal, Europe quickly took over as the largest source of exports starting in mid-2021 after transmission established there (Figure S1A). For several months following this, the number of Delta exports from Europe even exceeded the number of introductions of the variant into Europe (Figure S2).

We inferred the dispersal patterns of Omicron lineages BA.1 and BA.2 separately, given their genetic distinctiveness, until March 2022. Although both lineages were detected first in southern Africa at around the same time, their dynamics of global dissemination differed. Consistent with the first major Omicron

BA.1 waves occurring in southern Africa, we infer that the earliest exportation events of BA.1 originated from there during November and December 2021. Most Omicron BA.1 international lineage movement occurred within western Europe (>2,000) and we infer the timing of those transmissions to be centered around mid-January 2021 (Figures 1 and S1A). There was also considerable spread of BA.1 within the Americas (~500 exports) and from North America to Europe (>800 exports) during the same time period (Figures 1 and S1A). In comparison there were only 70 and 191 estimated exportation events from southern Africa to North America and western Europe, respectively. We estimate that Omicron BA.2 early exportations from southern Africa were to the Indian subcontinent and to Europe, also starting in late November 2021 (Figure S1A). Germany, India, and the UK were the three largest exporters of Omicron BA.2 overall, with 171, 170, and 100 inferred exports, respectively. Africa received approximately the same number of re-introductions of BA.1 and BA.2 as were originally exported from the continent (~54 and ~60 inferred exports versus ~198 and ~88 inferred re-introductions). We infer North America to be the major source of BA.1 and Asia of BA.2 (Figure 1). Crucially, this means that the continent where these lineages were first reported did not act as a major source of the VOC's global dissemination, particularly after the first couple of months of the duration of dissemination (Figures 1 and S1B). Although both lineages

emerged around the same time in late 2021,[8] international dispersal of BA.2 occurred, on average, later than that of BA.1 (Figure 1). This is expected given that, globally, BA.1 expanded and fueled large epidemic waves before BA.2 and that BA.2 partially evaded BA.1 immunity[20] and potentially had a competitive advantage only after BA.1 had spread. Further, some countries were still experiencing large Delta waves when BA.1 and BA.2 were imported, possibly slowing their spread (e.g., Germany).

Exploring the mechanisms of variant expansion globally, we observe a small positive effect of the nationally recorded case incidence attributed to each variant on the inferred volumes of exports out of corresponding countries (Figure S3A; Table S1), although this is partly influenced by the case-sensitive genomic sampling strategy. We also observe a positive relationship between the connectivity of countries to global air travel networks and their inferred contribution to export numbers for all variants (Figure S3A). However, the effect of air travel passenger volume had smaller effects on viral export for Beta and Gamma in comparison with the other variants (Table S1). This effect is likely a combination of lower international travel during the time of dispersal of Beta and Gamma and their emergence in countries not on the backbone of international flight networks, whereas Alpha first emerged in a highly connected region. In fact, from December 2020 to March 2021, the total number of passengers out of the UK, South Africa, and Brazil was ~3.8 million, ~440,000 and ~860,000, respectively, adding to the understanding of a larger geographical reach of Alpha compared with Beta and Gamma. As expected, this suggests that both case incidence and mobility play a role in determining the extent to which a country participates in the global dissemination of variants, unless there is co-circulation of different variants in the world. In this case, local variant-specific incidence will be a better predictor of estimated variant exports from a certain country, and mobility will only matter in regions where a variant is dominant (Figure S3A).

Quantifying regional and global dissemination hubs

Next, we investigated global and regional viral exports to identify hubs of international dissemination of different VOCs. Our results reveal that the US was the largest contributor to VOC lineage exportations globally, responsible for ~30% of all inferred VOC exports to other countries, followed by India, the UK, South Africa, and Germany, which contributed roughly 20%, 12%, 6%, and 5% of global VOC exports, respectively (Figure 2). The share of contributions to international exportations is highly correlated with countries' total air travel passenger volume (Spearman correlation ρ = 0.71, p < 0.001, Figure S3D). However, we show that the role of important global hubs varied among VOCs. For instance, South Africa acted as a major global hub for viral exportation for the Beta variant. The US's role was most visible for Omicron BA.1 (~75% of US global exports), whereas India's share of global exports was dominated by Delta and Omicron BA.2 (~70% of India's global exports), and the UK's by Alpha (~48% of the UK's global exports). As with the dispersal patterns discussed above, the most important inferred contributors to Omicron dissemination globally were more proximal (from secondary locations) than distal (from southern Africa). These findings are well supported by reported epidemiological trends.

For example, the US experienced large BA.1 wave in late November and December 2021 in metropolitan and highly connected cities on the East Coast (Washington, D.C., and New York City). Similarly, a large number of BA.2 infections were reported from India after it had spread there from southern Africa.[21] In fact, because recorded infection numbers were much higher during Omicron waves (roughly 130 million BA.1 and 110 million BA.2 infections globally within 5 months of circulation) compared with prior VOC waves (e.g., roughly 20 million Alpha and 90 million Delta infections globally in 10 and >12 months of circulation, respectively) due to higher transmissibility and relaxed restrictions, we generally infer a higher contribution of global and regional hubs toward the dissemination of Omicron BA.1 and BA.2 compared with, for instance, Alpha or Delta, respectively. This dynamic is substantiated by the greater effect of air travel passenger volume on viral exportation for Omicron BA.1 and BA.2, during a time of higher global connectivity, when compared with the effect of this variable on exports of the other variants (Table S1). We observe that countries that acted as major global hubs were also important in disseminating VOCs regionally (i.e., within the same continent). For example, Beta and Gamma were variants that mainly expanded regionally, and we find that the dissemination of these variants in Africa and South America accounted for >50% of exportations in those regions (Figure 2). A few countries also emerge as large regional hubs of viral exportations despite low global contributions: for example, the Philippines and Pakistan in Asia, Colombia in South America, and Spain and Italy in Europe (Figure 2), potentially linked to a combination of early seeding, large epidemics, and passenger numbers.

The inferred networks of viral dispersal that we discuss here are dynamic, as shown by the heterogeneous patterns when comparing VOCs. These are likely influenced by localized epidemic sizes, human mobility, population immunity, and viral variant phenotypes. Air travel passenger volume was estimated to have a statistically significant increase in viral exportations for all variants, with the greatest effect for BA.1 and BA.2 (Table S1). The number of local cases and deaths for each country also had a significant, although very small, positive relationship with viral exports. Although mixed effects were found with the influence of international travel bans. Generally, the implementation of stricter travel bans, for example, level 4 of total border closures, reduced viral exports (except for Alpha) when compared with level 1 (screening of passengers). These results were not statistically significant for Omicron BA.1 and BA.2, likely because global circulation of these variants was already occurring before the implementation of these bans on countries and regions of presumed origin. An immediate implication of these findings is that the tendency of countries to act as important global or regional hubs of viral dissemination is dictated by local case incidence and the countries' position in the hierarchical global air travel network,[22] which we explore further in the last section of the results, rather than the location of emergence of variants (Figure S3A; Table S1).

Role of first-reporting countries in global VOC dispersal

Following the emergence of each VOC, a variety of travel and passenger quarantine restrictions were put in place.[23–25] For
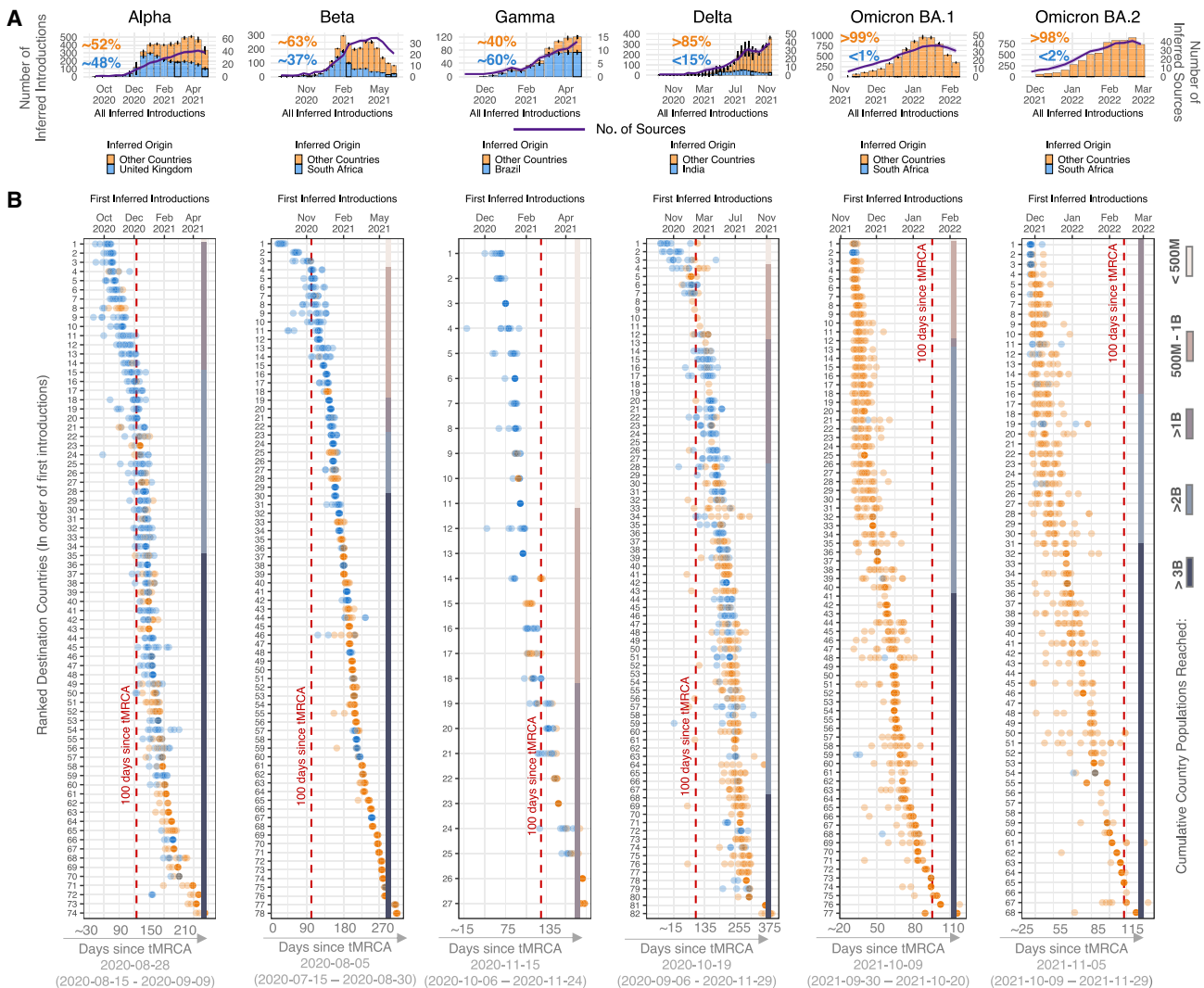
Figure 3. Inferred origins of global VOC dissemination events

(A) Changes in proportions of all inferred introductions from the country of presumed origin for each VOC (bars) and the number of countries inferred to be acting as onward sources of each VOC (purple line, with scale in the second y axis). Results shown are determined from 10 replicates of genome subsampling. Error bars indicate standard deviation.

(B) Date of first inferred introduction per country, shown as circles, colored by location of origin, i.e., presumed origin (blue) or not (orange). The y axis represents countries, which are ranked and labeled by the median of their dates of first introduction (from 10 replicates). The lower x axis denotes the delay between the estimated median TMRCA (with confidence interval range dates shown for each VOC, as reported in published studies[1–3,6,8,26]). The bars on the right side of each panel represent the cumulative population where the variant has been reported, calculated as the sum of the country populations that have observed introductions up to that point.

example, travel restrictions were implemented on travel from South Africa for 291 days after the discovery of Beta, on travel from South America for 270 days following the discovery of Gamma, and then again on South Africa for around a month for Omicron.[23] Here, we investigate the inferred sources and timing of international introductions of VOCs from the place of first detection and contrast them to importations from all other locations (UK for Alpha, South Africa for Beta, Omicron BA.1 and BA.2, Brazil for Gamma, and India for Delta).

The main finding for all VOCs is that even though VOC exports initially occurred mostly from the country of first reporting or pre-

sumed origin, this progressively shifted, and more countries became sources of exports to other locations as incidence in other countries increased (Figure 3). We observe that this shift happened much more rapidly for Delta and Omicron compared with Alpha, Beta, and Gamma (Figure 3). For Alpha, we find that by December 2020, roughly 100 days after the estimated date of emergence of this variant (time to most common recent ancestor [TMRCA]), around 20 countries, totaling over 2 billion population, had already received at least one inferred introduction of the variant (an underestimate of the real number of introductions) and were themselves acting as sources of

exportations from established local transmission (Figure 3B). This suggests wide geographic dispersal and cryptic transmission by the time travel restrictions were implemented around December 2020 to control dissemination of Alpha. This was seen by the positive effect of international travel bans on viral exports for Alpha (Table S1), whereas travel bans were found to reduce viral exports for the other variants. This pattern of gradual shift of the inferred source away from the presumed country of origin is similar for Beta and Gamma, the only difference being the lower number of countries, cumulating less than 1 billion population, inferred to have received an introduction of Gamma. Of all introductions to other locations that we infer, the UK contributed 48% of Alpha, South Africa contributed 37% of Beta, and Brazil contributed 60% of Gamma (Figure 3A).

The pattern is different from Delta and Omicron BA.1 and BA.2 lineages: India and South Africa, the presumed countries of origin of Delta and Omicron, respectively, very rapidly transitioned to being minor sources of both the first and overall introductions of these VOCs to other countries. In the case of Delta, fewer than 15% of all introductions to other countries were attributed to India (Figure 3A). Although India's contribution to first and overall Delta introductions did not completely subside over time, it decreased due to the contribution of other countries as a source of Delta. The shift away from the presumed location of origin is even more marked for Omicron BA.1 and BA.2. Overall, we infer that South Africa was the source of fewer than 1% and 2% of BA.1 and BA.2 introductions globally (Figure 3A). We show that within the first week (early November 2021) of BA.1 and BA.2 global dissemination, the first introductions to other countries were already originating from places other than South Africa. We also observe from the temporal reconstructions of these events that by the time travel restrictions were placed against southern Africa in December 2021, Omicron BA.1 exports could already be inferred from more than 30 other countries, cumulating over 2 billion population. Additionally, 100 days after the estimated TMRCA of Omicron BA.1 and BA.2, we could already infer introductions into more than 80 and 60 countries, respectively, in stark contrast to the dispersal of the other VOCs during the same time frame (Figure 3B).

For all the variants investigated here, the results point to the diminishing importance through time to the international dispersal of VOCs from the first presumed origin. This shift was more rapid with Delta and even more so with Omicron, potentially due to increased transmissibility of these variants, as well as fewer restrictions on travel and fewer PHSMs in many places, meaning higher and more sustained local transmission and thus more opportunity for onward spread. These conditions allowed these VOCs to reach other countries more rapidly, often even before first detection of the variants by virus genome sequencing and in all cases before any travel restrictions were implemented. The fact that this phenomenon is even more notable for Omicron could further be explained by a considerable increase in international travel volumes at the end of 2021, whereas travel bans against southern African countries were implemented very rapidly following the first report of Omicron emergence. These findings must be considered in light of the sensitivity of inferring VOC importation origins using genomic data alone, which can be impacted by sampling bias, especially during the earliest phase of VOC emergence. Inferences can be improved by using independent data sources (e.g., estimated importation intensity [EII] presented in McCrone et al.[6]) or by integrating individual travel histories from genomic sequence data.[27]

Impact of international travel on VOC dispersal

The transmission of SARS-CoV-2 was accompanied by major shifts in human mobility patterns throughout the pandemic.[10,15,28] In addition to national lockdowns restricting local movements and mixing, varying levels of air passenger travel restrictions were implemented in response to the initial emergence of SARS-CoV-2 and subsequent waves of transmission. In reconstructing the global dispersal of VOCs, the substantial decreases and more recent increases in international air travel might explain the variation in dispersal of VOCs alongside differences in immunity and vaccination. To examine how air travel has influenced the speed of dissemination of VOCs worldwide, we investigated global air travel passenger volumes between February 2020 and March 2022 and the network structure of the global airline network and compared them with the speed of dispersal of VOCs in countries reporting VOCs using genomic data. This delay was quantified here as the number of days between the TMRCAs (median) of each VOC (Omicron BA.1 and BA.2 separately) from published studies (Table S2) and the date of collection of the first sequenced VOC sample in each country (source: GISAID).

We find that, among countries, the Alpha, Beta, and Gamma variants were first sampled for sequencing on average 64–425, 95–300, and 48–251 days (5th–95th percentiles), respectively, after their emergence (Figure 4A, network visualization for Alpha shown in Figure S4A). On average, it took longer for countries to first sequence the Delta variant, with a delay of 45–336 days (5th–95th percentiles) after its estimated date of emergence in October 2020[6] (Figure 4A, network visualization for Delta shown in Figure S4B). The relatively longer delay between emergence and dates of sequencing can be potentially explained by the rapid spread of the Alpha, Beta, and Gamma variants during that time in other countries and the relatively longer period between emergence and rapid spread of Delta in India prior to global dissemination.[7] In the case of Omicron, both the BA.1 and BA.2 lineages dispersed around the world much faster than did the preceding VOCs. Omicron BA.1 and BA.2 were sampled on average just 7–98 days and 28–186 days (5th–95th percentiles), respectively, after their emergence (Figure 4A). This was likely strongly influenced by a 3-fold increase in global air travel passenger volumes, to ~60 million per month during the spread of Omicron, compared with ~20 million per month a year before during the time when the Alpha, Beta, Gamma, and Delta variants disseminated (Figure 4A). Additionally, it is now well known that Omicron lineages were highly immune evasive, causing infections globally at much higher rates (see also next section).

To explore more specifically the relationship between global air travel and the velocity of global VOC dispersal, we calculated pairwise correlations between sampling delays for each country for each VOC against a measure of distance along the most probable path connecting two countries (hereafter referred to as effective distance, D_{eff}), the total incoming travel volumes to that country during the time of VOC dispersal, and incoming travel
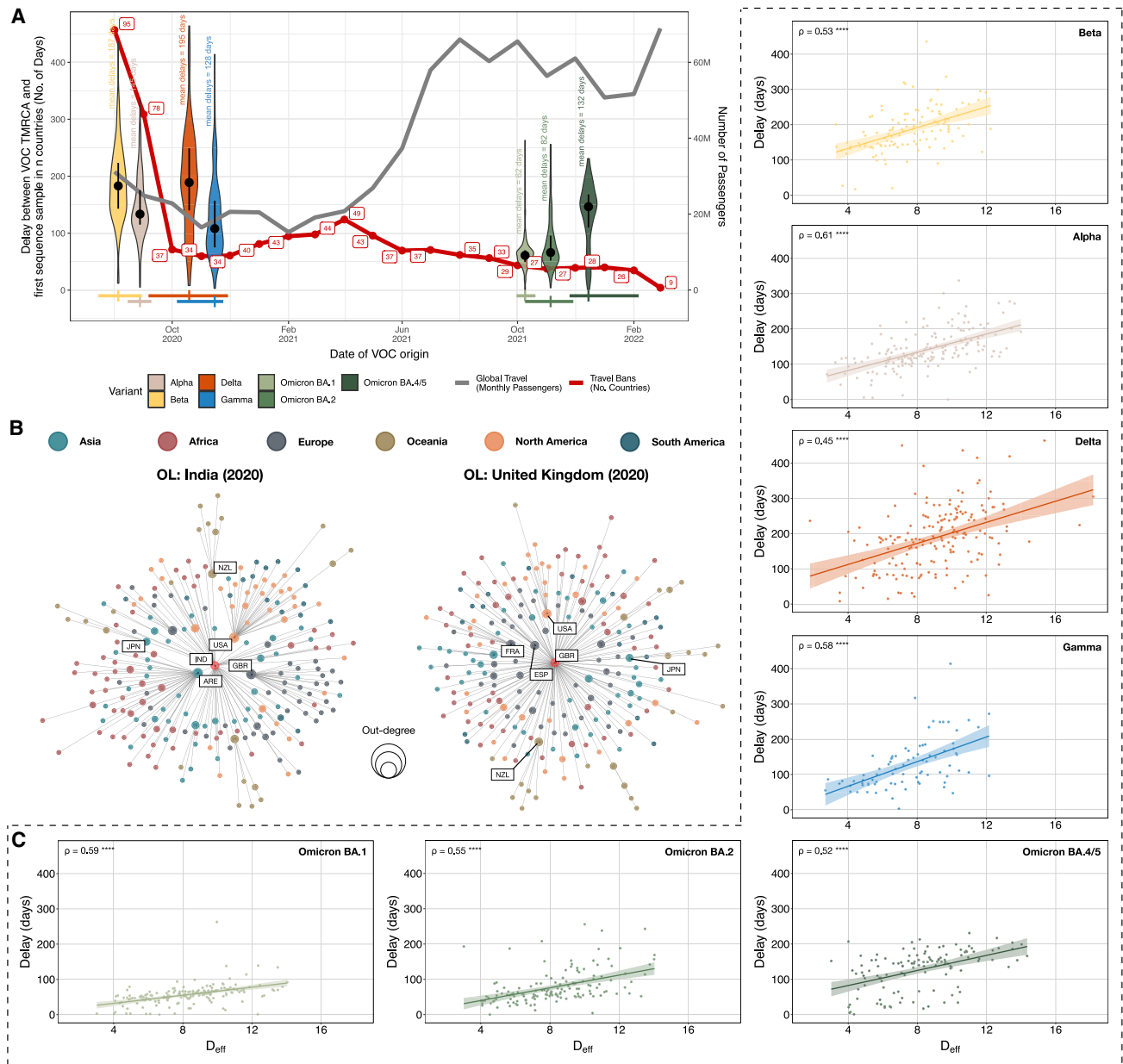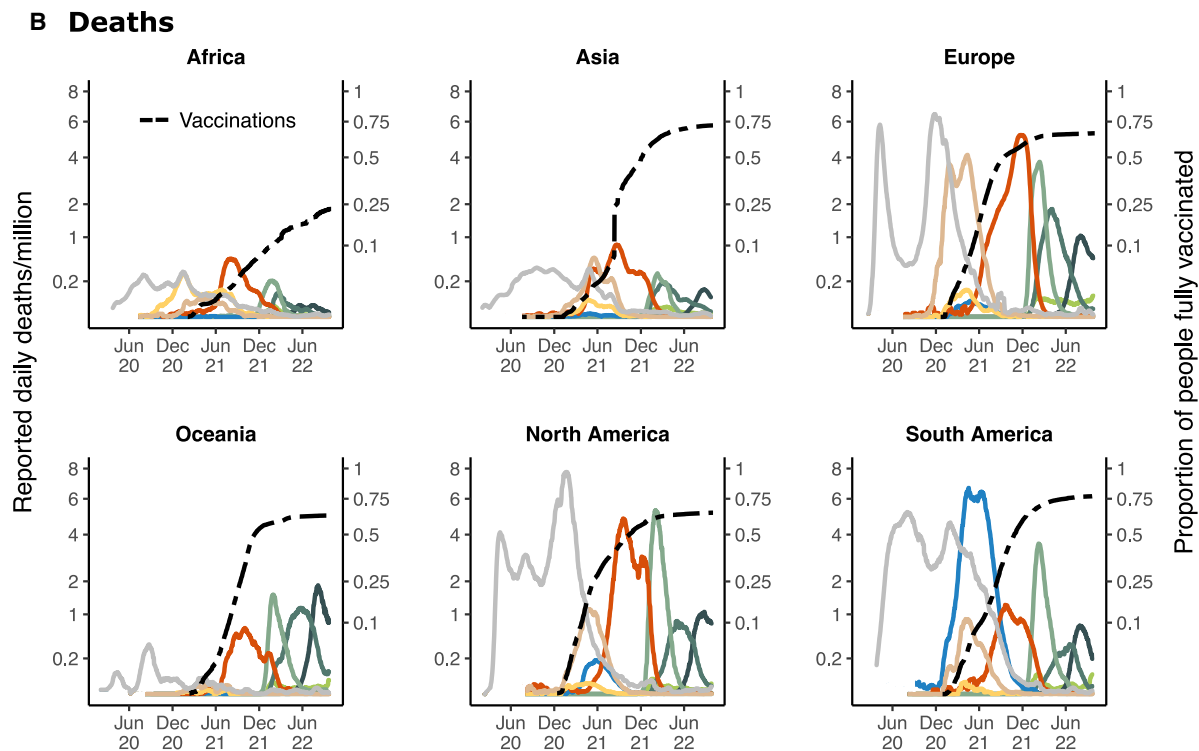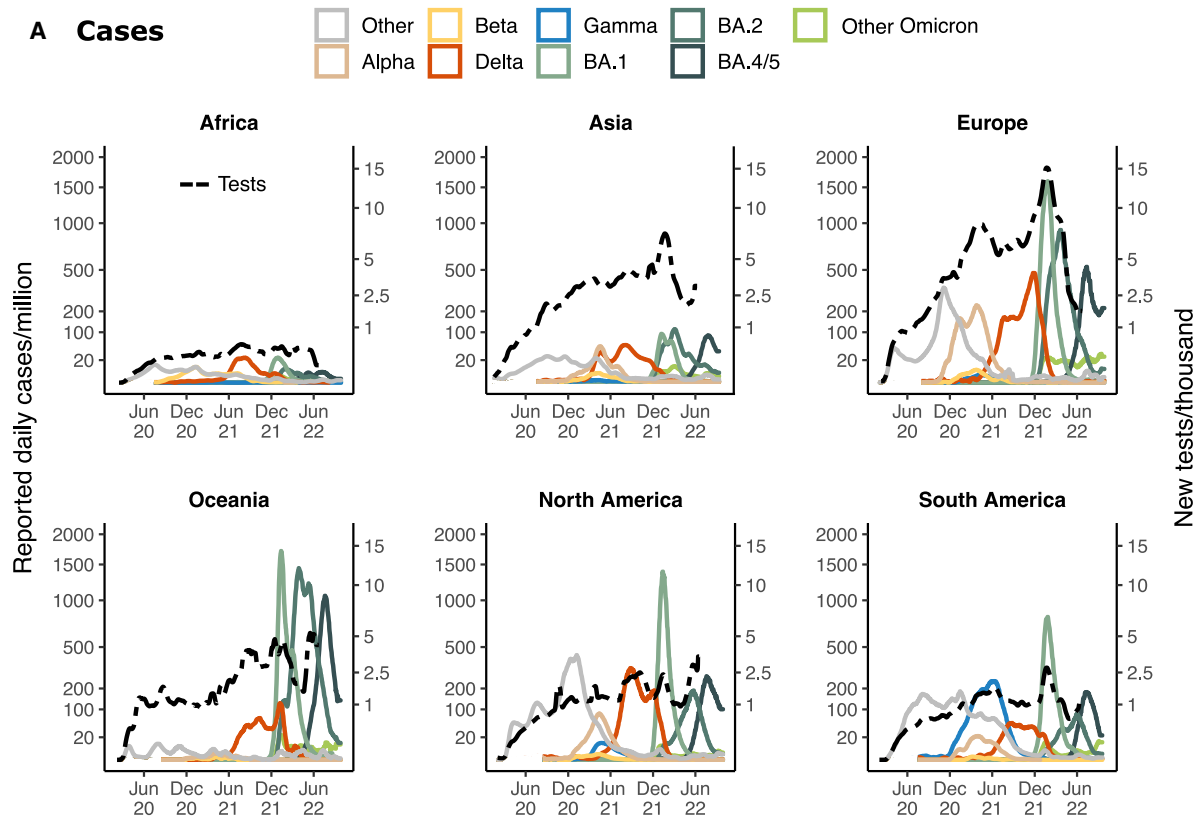
Figure 4. Impact of global air travel on VOC dissemination

(A) Delay (number of days since TMRCA) of each VOC to be first sampled in countries around the world, total global monthly air passenger volumes from September 2020 to March 2022, and the number of countries with active travel bans in the same period (data are sourced from the Oxford COVID-19 Government Response Tracker project [https://github.com/OxCGRT/covid-policy-tracker], where countries with international travel controls of levels 3 or 4 were counted as having travel bans in place). The corresponding mean of each violin plot is shown. The dot and error bars inside each group denote the median and interquartile range, respectively. Dates of VOC origin are taken as their published mean estimated dates of emergence (TMRCA), with crosses representing the median and high confidence range values.[1–3,6,8,26] The date of arrival of each VOC per country is taken to be the first sampling date of a sequenced case in GISAID (date of access: 18 September 2022).

(B) The shortest path tree constructed using global air traffic data from Oct-2020, with India (left) and the United Kingdom (right) as the presumed origin location (OL). Each node represents a country and is colored according to the continent. The radial distance of each node from the presumed OL along the connecting branches represents the effective distance D_{eff}.[22] The radius of each node scales with the number of descendant nodes (out-degree).

(C) Scatterplot and Spearman's rank correlation coefficient of the effective distance D_{eff} against delay in first sampling of VOCs in countries globally. The correlation coefficient is indicated for each VOC, with the level of significance indicated by the number of asterisks. A best-fit line is shown for each VOC, with the shaded band indicating 95% confidence interval.

A Cases

Africa · Asia · Europe · Oceania · North America · South America

B Deaths

Africa · Asia · Europe · Oceania · North America · South America

(legend on next page)

volumes from the presumed origin country only for each VOC. The effective distance (see method details) is a measure of how likely a randomly chosen individual in an origin country is to travel to another country via the most probable path given the underlying mobility network, which in our case is the global airline network (Figure 4B). We observe large differences in the airline network depending on the origin location, but rather surprisingly not between October 2020 and 2021 (Figure S5); for example, India (Figure 4B) connects to large hubs, which then connect to many countries, whereas the UK has a more star-like network structure with direct connections to many countries in the world. This is also reflected when comparing the effective distance, with passenger flows showing larger differences when considering India as the origin node vs. the UK (Figure S6A). Using this measure, we find that the delay of the first sample of VOCs is positively correlated with the effective distance (Figure 4C). The association between arrival times of VOCs and passenger volumes is comparable but slightly weaker (Table S3). However, for Omicron BA.1 and BA.4/5, the effective distance measure is able to better predict the arrival times as compared with using the raw travel volumes from South Africa (Spearman's rank correlation 0.47 vs. 0.59 and 0.39 vs. 0.52, respectively), pointing toward a larger contribution from major transit hubs in the global dissemination than in the case of Alpha, Delta, and Gamma. This further supports previously discussed findings that locations other than the presumed origin quickly became exporters of viral lineages for a highly transmissible variant. Interestingly, for the Delta variant, we observe a substantially longer than expected delay in first sampling in the United Arab Emirates (ARE), despite its relatively small effective distance from India (the presumed origin location of Delta) and that it acts as a transit node between the shortest paths from India to many countries in Asia and Africa (Figure 4B, left). This suggests that ARE potentially acted as an epidemiologically important yet undetected secondary hub in the global dissemination of the Delta variant, likely as a result of its very low sequencing intensity with only 2,630 sequences uploaded to GISAID since the beginning of the pandemic. To lend further support to this hypothesis, we performed a sensitivity analysis by removing ARE from the global air traffic network and recomputing the effective path from India to each country—although substantial changes to the shortest path tree are observed as a result of the removal of ARE, the predicted arrival times of Delta according to the new effective distances are not significantly different from those before ARE was removed (Figures S6B and S6C). The observed arrival times of Delta are therefore inconclusive as evidence of ARE being a major secondary transit node in the global dissemination of the variant—future work should consider the variance of the shortest path tree given the observed air traffic network and therefore the statistical confidence of the topological position of each node in the tree.

Global epidemic and variant dynamics

In addition to the distinct dispersal dynamics described above, VOCs emerged in globally heterogeneous epidemiological landscapes (Figure 5). Variations in epidemic intensity around the world were further exacerbated by uneven diagnostic testing rates (Figure 5A), distinct levels of population immunity as the pandemic progressed (either vaccine or infection acquired), and variation in geographical and temporal drivers of transmission. Despite underreporting of infection numbers,[29] combining death, genomic surveillance, and testing data provides qualitative insights into the differences in epidemic waves across continents (Figure 5). Although Alpha, Beta, and Gamma expanded regionally, Delta and Omicron swept across the globe, becoming dominant worldwide in mid-late 2021 and early 2022, respectively (Figure 5A).

Throughout the different waves of infection, we observe a marked difference in reported mortality in Africa, Asia, and Oceania compared with Europe and South and North America (Figure 5B). This is likely due to a combination of under-reported mortality in Africa and Asia, as suggested by high levels of modeled excess deaths,[29] and low virus circulation in Oceania due to prolonged border closures in earlier stages of the pandemic.[30,31] Despite high vaccination coverage and accelerated booster rollout during the emergence of BA.1 (Figure 5B), case incidence increased rapidly (Figure 5A) in the context of relaxing non-pharmaceutical interventions (NPIs) across the world. The high number of cases meant considerable mortality, especially in those few locations with no or low immunity, due to a combination of low population exposure to the virus and low vaccination rates in high-risk groups, for example, in Hong Kong.[32] In absolute numbers, Oceania and North and South America all experienced higher mortality due to Omicron BA.1 than Delta. Vaccine inequity exacerbated the inability to protect even high-risk groups in many places at this stage, particularly in the context of low vaccination coverage in Africa (Figure 5B).

DISCUSSION

Our study shows that SARS-CoV-2 VOCs (Alpha, Beta, Gamma, Delta, and Omicron BA.1 and BA.2) disseminated around the world according to different spatial source-sink dynamics and that global travel hubs were important contributors to viral exportations. The international spread of Delta and Omicron was substantially different from that of Alpha, Beta, and Gamma. The dispersal of Delta and Omicron was in general more multi-focal, with multiple regions contributing to their global invasion. Also, Australia contributed to viral exchanges for Delta and Omicron (unlike for Alpha, Beta, and Gamma), consistent with published reports of Delta introductions despite quarantine measures

Figure 5. Continental epidemiology of SARS-CoV-2 cases, mortality, testing, and vaccination

(A) The progression of daily reported cases per continent from February 2020 to October 2022 (log scale, first y axis). The 7-day rolling average of daily reported case numbers is colored by the inferred proportion of variants responsible for the infections, as calculated by genomic surveillance data (GISAID date of access: 1 October 2022) averaged over 20 days. The line shows the 7-day rolling average of the number of daily tests per thousand population per region (scale shown in the second y axis) aggregated for countries for which these data are available for each continent.

(B) The 7-day rolling average of daily reported deaths colored by the inferred proportion of variants, as calculated for case data, with an assumption of time lag of 20 days between infection and death applied (see more details in method details). The dashed line displays the proportion of people fully vaccinated per region (scale on second y axis), where those that received all doses prescribed by the initial vaccination protocol are considered fully vaccinated.

and the opening of borders for non-Australian citizens prior to the Omicron wave.[33–35] These differences reflect both the distinct global landscapes at different stages of the pandemic, and the intrinsic characteristics of different variants. Alpha, Beta, and Gamma circulated in more restricted sets of locations, whereas Delta and Omicron dominated infections in a global sweep. The period during which Alpha, Beta, and Gamma emerged was characterized by lower global mobility and widespread travel restrictions, whereas the gradual lifting of such restrictions was associated with the more rapid and widespread Delta and Omicron disseminations.[6]

We also investigated the role of the presumed origin location of each VOC (the countries that first reported each variant) in the global dispersal of these viruses. Although we infer that the UK, South Africa, and Brazil were the source of the majority of global exportations of Alpha, Beta, and Gamma (all >35%), we find that for Delta and Omicron, the contribution of the presumed origin location was much smaller (<15%). We observe differences in the speed at which countries that are not the presumed origins became exporters of the VOC. For Delta and Omicron, this rapid transition is attributed to a mix of increased transmissibility and increased global air travel. Our results should be viewed in the context of a country's epidemiological landscape. The pattern that we present, however, highlights that locations with high case incidence and global connectivity have the potential to become major contributors to variant exportations if early seeding of viral variant outbreaks is not controlled.

SARS-CoV-2 variants continue to emerge and spread worldwide, as seen most recently in the case of the Omicron BA.4 and BA.5 lineages[26] and sub-lineages. In this context, our findings have some general implications for public health. First, we show that once a variant has been established in multiple countries, continued international spread may almost be inevitable. When specific routes are closed due to travel restrictions, other locations become responsible for a greater share of global dissemination. This indicates that in the case of emerging variants, especially those with enhanced virulence and waning immunity, actions to control or mediate the effects of virus transmission should be undertaken everywhere. Second, our results indicate that travel restrictions, especially targeted ones, are often implemented after initial imports have already come into other countries, especially for more transmissible variants, as discussed elsewhere.[25,36] To limit the extent of local transmission, a combination of measures is necessary, including testing at arrival, antigen testing before large gatherings, isolation while infectious, and vaccination, among others.[37] Lastly, as global air travel and human mobility return to pre-pandemic levels and beyond, new variants are likely to reach secondary countries much faster, potentially before being identified by genomic surveillance. Despite the massive effort of genomic sequencing globally, the nature of respiratory viruses such as SARS-CoV-2 and especially highly transmissible VOCs means that testing and genomic surveillance will often struggle to detect a new variant and determine its significance before there is ample opportunity for wide dispersal.[38] This makes targeted travel restrictions increasingly ineffective, and continued investment and innovation in robust, fast, and systematic diagnostic and surveillance programs is crucial for current and future pathogens. For example, targeted genomic surveillance can be informed by a location's position in the hierarchical global air travel network.

Future work to document the extent of the impact of the global spread of VOCs could additionally consider local contexts in relation to VOC characteristics, including population immunity profiles, whether acquired through vaccination or through previous infection waves of particular variants, parameters of waning immunity related to the duration of time since the last wave, and local control measures. By systematically analyzing large representative datasets for each variant, this study highlights the role of global and regional hubs in viral dispersal while contrasting it with targeted travel restrictions toward the presumed origins of VOCs. Using travel data, we discuss that novel emerging VOCs with a clear transmission advantage are likely to spread much faster around the world given today's increase in travel volumes. Overall, the global-scale spatiotemporal invasion patterns described here provide an opportunity to integrate knowledge of viral exportation and importation dynamics into our collective understanding of pandemic progression. This will be critical both in managing the upcoming stages of global SARS-CoV-2 transmission and within preparedness plans for future epidemics.

Limitations of the study

The findings presented here are derived primarily from phylogeographic analysis and have several limitations. Our genomic sampling was informed by the timing and size of epidemics per country to avoid over-representing countries with more sequencing and to focus inferences on a subset of sequences that most accurately represents routine surveillance of local transmission rather than possible targeted sequencing of travel-related cases or over-sequencing at the beginning of waves (method details). However, we cannot rule out remaining biases in our datasets, particularly those associated with uneven testing rates and case reporting globally. With this method of sampling, it is also not possible to unambiguously identify the very first introductions of VOCs, especially if they are associated with sustained cryptic transmission before an increase in variant-related cases is noted. We also rely on national genomic surveillance data to scale reported cases to variants, and we recognize biases arising from that, including possible uneven representation of incidence from subregions of countries, and non-uniform sequencing proportions at various stages of the pandemic. Furthermore, the global and continental view taken in our analysis will obscure fine-scale epidemiological heterogeneity within countries. In order to best represent VOC epidemics worldwide, we performed phylogenetic inferences on datasets of roughly 20,000 genomes for each VOC, in replicates of 10. Given the size of these datasets, we were not able to employ full Bayesian phylogeographic reconstruction methods given the availability of computational resources. This would mostly affect the precise timing of inferred ancestral state changes per country; however, the very dense sampling of virus genomes through time puts some strong temporal constraints on when a state change can occur. Finally, our analysis focuses on the global spatiotemporal invasion of VOCs and does not attempt to study the impact on epidemic growth in countries that do receive an introduction of these VOCs. Therefore, we cannot make a causal claim between the size of the epidemic wave in the destination country and the

number of viral introductions and travel volumes. Other studies have shown that local measures in the destination country influence the control of waves more than the number of seeding events.[6]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead Contact
  - ○ Materials Availability
  - ○ Data and Code Availability
- METHOD DETAILS
  - ○ Epidemiological Data and Genomic Prevalence
  - ○ Genomic sampling
  - ○ Phylogenetic reconstruction
  - ○ Source sink dynamics and viral movement patterns quantification
  - ○ Air Travel Data
  - ○ Air travel network and sensitivity analysis
  - ○ Calculation of Deff
  - ○ Comparisons between D_{eff} and travel volumes
  - ○ Comparison between D_{eff} calculated using data from 2020 and 2021
  - ○ Sensitivity analysis of United Arab Emirates (ARE) as a major travel hub
  - ○ Regression of travel volume, travel restrictions, and case and death counts on viral exports
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.06.001.

AUTHOR CONTRIBUTIONS

Conceptualization, H.T., E.W., M.U.G.K., and T.d.O.; methodology and data analysis, H.T., E.W.., A.F.B., D.M., M.M., M.G., K.K, C.H., I.I.B., J.E.S., J.L.-H.T., J.P., J.S.X., D.d.S.C., and F.R.; supervision, N.R.F., M.U.G.K., and T.d.O.; writing – original draft: H.T., E.W., R.J.L., D.M., and M.U.G.K.; writing – review & editing: H.T., E.W., J.L.-H.T., M.M., M.U.G.K., C.B., O.G.P., T.d.O., R.J.L., and N.R.F.

DECLARATION OF INTERESTS

K.K. is the founder of BlueDot, a social enterprise that develops digital technologies for public health. C.H. is employed at BlueDot.

REFERENCES

1. Hill, V., Du Plessis, L., Peacock, T.P., Aggarwal, D., Colquhoun, R., Carabelli, A.M., Ellaby, N., Gallagher, E., Groves, N., Jackson, B., et al. (2022). The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. Virus Evol. 8, veac080. https://doi.org/10.1093/ve/veac080.

2. Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., et al. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. Nature 592, 438–443. https://doi.org/10.1038/s41586-021-03402-9.

3. Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.D.S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 372, 815–821. https://doi.org/10.1126/science.abh2644.

4. Tao, K., Tzou, P.L., Nouhin, J., Gupta, R.K., de Oliveira, T., Kosakovsky Pond, S.L., Fera, D., and Shafer, R.W. (2021). The biological and clinical significance of emerging SARS-CoV-2 variants. Nat. Rev. Genet. 22, 757–773. https://doi.org/10.1038/s41576-021-00408-x.

5. Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, Á., et al. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature. https://doi.org/10.1038/s41586-021-03470-x.

6. McCrone, J.T., Hill, V., Bajaj, S., Pena, R.E., Lambert, B.C., Inward, R., Bhatt, S., Volz, E., Ruis, C., Dellicour, S., et al. (2022). Context-specific emergence and growth of the SARS-CoV-2 Delta variant. Nature 610, 154–160. https://doi.org/10.1038/s41586-022-05200-3.

7. Dhar, M.S., Marwal, R., Vs, R., Ponnusamy, K., Jolly, B., Bhoyar, R.C., Sardana, V., Naushin, S., Rophina, M., Mellan, T.A., et al. (2021). Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. Science 374, 995–999. https://doi.org/10.1126/science.abj9932.

8. Viana, R., Moyo, S., Amoako, D.G., Tegally, H., Scheepers, C., Althaus, C.L., Anyaneji, U.J., Bester, P.A., Boni, M.F., Chand, M., et al. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature 603, 679–686. https://doi.org/10.1038/s41586-022-04411-y.

9. Brauner, J.M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A.B., Leech, G., Altman, G., Mikulik, V., et al. (2021). Inferring the effectiveness of government interventions against COVID-19. Science *371*. https://doi.org/10.1126/science.abd9338.

10. Tegally, H., Khan, K., Huber, C., de Oliveira, T., and Kraemer, M.U.G. (2022). Shifts in global mobility dictate the synchrony of SARS-CoV-2 epidemic waves. J. Travel Med. *29*. https://doi.org/10.1093/jtm/taac134.

11. Shu, Y., and McCauley, J. (2017). GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. *22*, 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

12. Hodcroft, E.B., Zuber, M., Nadeau, S., Vaughan, T.G., Crawford, K.H.D., Althaus, C.L., Reichmuth, M.L., Bowen, J.E., Walls, A.C., Corti, D., et al. (2021). Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Nature *595*, 707–712. https://doi.org/10.1038/s41586-021-03677-y.

13. Lemey, P., Ruktanonchai, N., Hong, S.L., Colizza, V., Poletto, C., Van den Broeck, F., Gill, M.S., Ji, X., Levasseur, A., Oude Munnink, B.B., et al. (2021). Untangling introductions and persistence in COVID-19 resurgence in Europe. Nature *595*, 713–717. https://doi.org/10.1038/s41586-021-03754-2.

14. Rito, T., Richards, M.B., Pala, M., Correia-Neves, M., and Soares, P.A. (2020). Phylogeography of 27,000 SARS-CoV-2 genomes: Europe as the major source of the COVID-19 pandemic. Microorganisms *8*. https://doi.org/10.3390/microorganisms8111678.

15. Kraemer, M.U.G., Hill, V., Ruis, C., Dellicour, S., Bajaj, S., McCrone, J.T., Baele, G., Parag, K.V., Battle, A.L., Gutierrez, B., et al. (2021). Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. Science *373*, 889–895. https://doi.org/10.1126/science.abj0113.

16. Wilkinson, E., Giovanetti, M., Tegally, H., San, J.E., Lessells, R., Cuadros, D., Martin, D.P., Rasmussen, D.A., Zekri, A.-R.N., Sangare, A.K., et al. (2021). A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. Science *374*, 423–431. https://doi.org/10.1126/science.abj4336.

17. Tegally, H., San, J.E., Cotten, M., Moir, M., Tegomoh, B., Mboowa, G., Martin, D.P., Baxter, C., Lambisia, A.W., Diallo, A., et al. (2022). The evolving SARS-CoV-2 epidemic in Africa: insights from rapidly expanding genomic surveillance. Science *378*, eabq5358. https://doi.org/10.1126/science.abq5358.

18. Brito, A.F., Semenova, E., Dudas, G., Hassler, G.W., Kalinich, C.C., Kraemer, M.U.G., Ho, J., Tegally, H., Githinji, G., Agoti, C.N., et al. (2022). Global disparities in SARS-CoV-2 genomic surveillance. Nat. Commun. *13*, 7003. https://doi.org/10.1038/s41467-022-33713-y.

19. Rahaman, M.M., Sarkar, M.M.H., Rahman, M.S., Islam, M.R., Islam, I., Saha, O., Akter, S., Banu, T.A., Jahan, I., Habib, M.A., et al. (2022). Genomic characterization of the dominating Beta, V2 variant carrying vaccinated (Oxford-AstraZeneca) and nonvaccinated COVID-19 patient samples in Bangladesh: A metagenomics and whole-genome approach. J. Med. Virol. *94*, 1670–1688. https://doi.org/10.1002/jmv.27537.

20. Zou, J., Kurhade, C., Xia, H., Liu, M., Xie, X., Ren, P., and Shi, P.Y. (2022). Cross-neutralization of omicron BA.1 against BA.2 and BA.3 SARS-CoV-2. Nat. Commun. *13*, 2956. https://doi.org/10.1038/s41467-022-30580-5.

21. Dhawan, M., Priyanka, and Choudhary, O.P. (2022). Emergence of Omicron sub-variant BA.2: is it a matter of concern amid the COVID-19 pandemic? Int. J. Surg. *99*, 106581. https://doi.org/10.1016/j.ijsu.2022.106581.

22. Brockmann, D., and Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. Science *342*, 1337–1342. https://doi.org/10.1126/science.1245200.

23. Mendelson, M., Venter, F., Moshabela, M., Gray, G., Blumberg, L., de Oliveira, T., and Madhi, S.A. (2021). The political theatre of the UK's travel ban on South Africa. Lancet *398*, 2211–2213. https://doi.org/10.1016/S0140-6736(21)02752-5.

24. Schermerhorn, J., Case, A., Graeden, E., Kerr, J., Moore, M., Robinson-Marshall, S., Wallace, T., Woodrow, E., and Katz, R. (2022). Fifteen days in December: capture and analysis of Omicron-related travel restrictions. BMJ Glob. Health *7*. https://doi.org/10.1136/bmjgh-2022-008642.

25. Kucharski, A.J., Jit, M., Logan, J.G., Cotten, M., Clifford, S., Quilty, B.J., Russell, T.W., Peeling, R.W., Antonio, M., and Heymann, D.L. (2022). Travel measures in the SARS-CoV-2 variant era need clear objectives. Lancet *399*, 1367–1369. https://doi.org/10.1016/S0140-6736(22)00366-X.

26. Tegally, H., Moir, M., Everatt, J., Giovanetti, M., Scheepers, C., Wilkinson, E., Subramoney, K., Makatini, Z., Moyo, S., Amoako, D.G., et al. (2022). Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. Nat. Med. *28*, 1785–1790. https://doi.org/10.1038/s41591-022-01911-2.

27. Lemey, P., Hong, S.L., Hill, V., Baele, G., Poletto, C., Colizza, V., O'Toole, Á., McCrone, J.T., Andersen, K.G., Worobey, M., et al. (2020). Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. Nat. Commun. *11*, 5110. https://doi.org/10.1038/s41467-020-18877-9.

28. Schlosser, F., Maier, B.F., Jack, O., Hinrichs, D., Zachariae, A., and Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. Proc. Natl. Acad. Sci. USA *117*, 32883–32890. https://doi.org/10.1073/pnas.2012326117.

29. Wang, H., Paulson, K.R., Pease, S.A., Watson, S., Comfort, H., Zheng, P., Aravkin, A.Y., Bisignano, C., Barber, R.M., Alam, T., et al. (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. Lancet *399*, 1513–1536. https://doi.org/10.1016/S0140-6736(21)02796-3.

30. Adekunle, A., Meehan, M., Rojas-Alvarez, D., Trauer, J., and McBryde, E. (2020). Delaying the COVID-19 epidemic in Australia: evaluating the effectiveness of international travel bans. Aust. N. Z. J. Public Health *44*, 257–259. https://doi.org/10.1111/1753-6405.13016.

31. Stobart, A., and Duckett, S. (2022). Australia's response to COVID-19. Health Econ. Policy Law *17*, 95–106. https://doi.org/10.1017/S1744133121000244.

32. Mefsin, Y.M., Chen, D., Bond, H.S., Lin, Y., Cheung, J.K., Wong, J.Y., Ali, S.T., Lau, E.H.Y., Wu, P., Leung, G.M., et al. (2022). Epidemiology of infections with SARS-CoV-2 omicron BA.2 Variant, Hong Kong, January-March 2022. Emerging Infect. Dis. *28*, 1856–1858. https://doi.org/10.3201/eid2809.220613.

33. Jelley, L., Douglas, J., Ren, X., Winter, D., McNeill, A., Huang, S., French, N., Welch, D., Hadfield, J., de Ligt, J., et al. (2022). Genomic epidemiology of Delta SARS-CoV-2 during transition from elimination to suppression in Aotearoa New Zealand. Nat. Commun. *13*, 4035. https://doi.org/10.1038/s41467-022-31784-5.

34. Zachreson, C., Shearer, F.M., Price, D.J., Lydeamore, M.J., McVernon, J., McCaw, J., and Geard, N. (2022). COVID-19 in low-tolerance border quarantine systems: impact of the Delta variant of SARS-CoV-2. Sci. Adv. *8*, eabm3624. https://doi.org/10.1126/sciadv.abm3624.

35. Milne, G.J., and Carrivick, J. (2022). SARS-CoV-2 Omicron disease burden in Australia following border reopening: a modelling analysis. Preprint at medRxiv. https://doi.org/10.1101/2022.03.09.22272170

36. Han, A.X., Kozanli, E., Koopsen, J., Vennema, H., RIVM COVID-19 molecular epidemiology group, Hajji, K., Kroneman, A., van Walle, I., Klinkenberg, D., Wallinga, J., et al. (2022). Regional importation and asymmetric within-country spread of SARS-CoV-2 variants of concern in the Netherlands. eLife *11*. https://doi.org/10.7554/eLife.78770.

37. Boyer, C.B., Rumpler, E., Kissler, S.M., and Lipsitch, M. (2022). Infectious disease dynamics and restrictions on social gathering size. Epidemics *40*, 100620. https://doi.org/10.1016/j.epidem.2022.100620.

38. Han, A.X., Toporowski, A., Sacks, J.A., Perkins, M.D., Briand, S., van Kerkhove, M., Hannay, E., Carmona, S., Rodriguez, B., Parker, E., et al. (2022).

Low testing rates limit the ability of genomic surveillance programs to monitor SARS-CoV-2 variants: a mathematical modelling study. Preprint at medRxiv. https://doi.org/10.1101/2022.05.20.22275319

39. Alpert, T., Brito, A.F., Lasek-Nesselquist, E., Rothman, J., Valesano, A.L., MacKay, M.J., Petrone, M.E., Breban, M.I., Watkins, A.E., Vogels, C.B.F., et al. (2021). Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell *184*, 2595–2604.e13. https://doi.org/10.1016/j.cell.2021.03.061.

40. Martin, D.P., Varsani, A., Roumagnac, P., Botha, G., Maslamoney, S., Schwab, T., Kelz, Z., Kumar, V., and Murrell, B. (2021). RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol. *7*, veaa087. https://doi.org/10.1093/ve/veaa087.

41. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 — approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490. https://doi.org/10.1371/journal.pone.0009490.

42. Sagulenko, P., Puller, V., and Neher, R.A. (2018). TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. *4*, vex042. https://doi.org/10.1093/ve/vex042.

43. Popescu, A.A., Huber, K.T., and Paradis, E. (2012). ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. Bioinformatics *28*, 1536–1537. https://doi.org/10.1093/bioinformatics/bts184.

44. Wickham, H. (2011). ggplot2. WIREs Comp. Stat. *3*, 180–185. https://doi.org/10.1002/wics.147.

45. Signorell, A. (2023). DescTools: tools for descriptive statistics. [R package DescTools version 0.99.48]. https://CRAN.R-project.org/package=DescTools.

46. Ward, T., and Johnsen, A. (2021). Understanding an evolving pandemic: an analysis of the clinical time delay distributions of COVID-19 in the United Kingdom. PLoS One *16*, e0257978. https://doi.org/10.1371/journal.pone.0257978.

47. Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat. Hum. Behav. *5*, 529–538. https://doi.org/10.1038/s41562-021-01079-8.

48. Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S, Fourth Edition (Springer). https://www.stats.ox.ac.uk/pub/MASS4/.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
SARS-CoV-2 Genomes	GISAID (GISAID: EPI_SET_230221dt)	https://www.gisaid.org/
International COVID-19 cases, deaths and vaccinations	Our World in Data	https://github.com/owid/covid-19-data/tree/master/public/data
Air passenger volumes (commercial)	IATA	https://www.iata.org/pages/default.aspx
International travel restrictions	Oxford COVID-19 Government Response Tracker	https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker
Software and Algorithms		
R	CRAN	https://cran.r-project.org/
subsampler	Alpert et al.[39]	https://github.com/andersonbrito/subsampler
RDP5	Martin et al.[40]	N/A
NextAlign	N/A	https://github.com/neherlab/nextalign
FastTree	Price et al.[41]	N/A
TreeTime	Sagulenko et al.[42]	https://github.com/neherlab/treetime
Ape R package	Popescu et al.[43]	CRAN
ggplot2 R package	Wickham[44]	CRAN
MASS R package	N/A	https://www.stats.ox.ac.uk/pub/MASS4/
DescTools R package	Signorell[45]	N/A

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for data and resources should be directed to and will be fulfilled by the Lead Contact, Houriiyah Tegally (houriiyah.tegally@gmail.com).

Materials Availability
This study did not generate new unique reagents, but raw data and code generated as part of this research can be found on public resources as specified in the data and code availability section below.

Data and Code Availability
The findings of this study are based on sequences and metadata associated with a total of 514,831 sequences collected in 141 countries and territories available on GISAID up to November 19, 2022, via gisaid.org (GISAID: EPI_SET_230221dt). All genome sequences and associated metadata in this dataset are published in GISAID's EpiCoV database. To view the contributors of each individual sequence with details such as accession number, Virus name, Collection date, Originating Lab and Submitting Lab and the list of Authors, visit https://doi.org/10.55876/gis8.230221dt. Custom data sources and scripts to reproduce the results of this study are publicly shared on GitHub (https://github.com/CERI-KRISP/SARS_CoV_2_VOC_dissemination). The repository contains all of the time scaled ML tree topologies, annotated tree topologies as well as custom data analysis and visualization scripts. Other datasets and pipelines used in this study are openly available and described in the method details section.

METHOD DETAILS

Epidemiological Data and Genomic Prevalence
We analyzed trends in daily numbers of cases of SARS-CoV-2 and reported deaths for each continent up to 1 October 2022 from publicly released data provided by the Our World in Data repository (https://github.com/owid/covid-19-data/tree/master/public/data). To provide a comparable view of epidemiological dynamics over time in different continents, the variable under primary consideration for Figure S13 was 'new cases per million' and 'new deaths per million'. Genomic metadata was downloaded for all entries on

GISAID for the same time period (date of access: 1 October 2022). From this, information extracted from all entries for this study included: date of sampling, continent of sampling, viral lineage and clade.

We calculate the rolling average of daily case and deaths numbers for each variant by inferring the daily proportion of variants responsible for infections, as calculated by genomic surveillance data on GISAID. A lag time of 17.4 to 24.7 days (median = 20) was applied between the calculated genomic prevalence and recorded deaths, as reported in published literature.[46] A smoothing factor of 20 was used to calculate the rolling averages for case numbers, but the smoothing factor used of death numbers varied according to the consistency of the data, as follows: Africa k = 20, Asia k = 20, Europe k = 14, Oceania k = 22, North America k = 14, South America k = 22.

Vaccination statistics per continent and testing and positivity datasets per country were also obtained from Our World in Data. Overall continental data is not available in the dataset for the testing and positivity rate, and was therefore calculated from countries for each continent with available data. The proportion of people fully vaccinated per region was calculated with the number of people fully vaccinated, defined as having received all doses prescribed by the initial vaccination protocol, and the population size data provided by Our World in Data.

Genomic sampling

Due to the sensitivity of ancestral state reconstruction methods to sampling bias, we performed our selection of genomic sequences in a careful manner to minimize biases to an extent that sampled datasets broadly reflect global reported case counts. From the complete set of entries for each VOC available on GISAID (Date of access: 17 September 2022), our subsampling strategy selected sequences to correspond to timeframes of global circulation of the respective VOCs, and in proportion to recorded cases in different countries. We used a previously described method *subsampler* (https://github.com/andersonbrito/subsampler)[39] to produce such globally representative subsets for each VOC. This method subsamples sequences per country based on case counts over the study period to ensure the random sample is both geographically, temporally and epidemiologically representative. In short, *subsampler* requires the sequence metadata for the complete dataset from which the subsampling occurs, along with a case count matrix, which we scaled for each VOC to their estimated prevalences for different countries based on GISAID data. This subsampling scheme was performed ten times using ten unique random number seeds for each VOC to produce ten randomly sampled genomic datasets per VOC. Subsampling within this scheme is performed using a baseline function, which represents the proportion of cases that the user wishes to sample. This was changed accordingly to produce datasets of approximately 20,000 sequences for each VOC. It is important to note here that due to country-specific sequencing efforts, some countries may not have sequenced enough cases to reach the desired sampling proportion stated by the baseline function. These countries are flagged as undersampled by showing negative values in the corresponding weekly sampling bias output file produced by this software. As a sensitivity analysis, we also performed this subsampling step using recorded COVID-19 deaths rather than case counts for Alpha, Beta and Gamma, to ascertain potential biases of testing rates in the resulting sampling proportions. We found that the sampling proportions for each country remains consistent whichever strategy is used and opt to perform the rest of the analysis with case sampling (Figures S7A–S7C). Due to the genetic distance between the BA.1 and BA.2 Omicron variants, these two sub-lineages were split in downstream analyses and in this subsampling procedure. To minimize the potentially confounding effects of recombination on downstream phylogenetics-based analyses (which assume sequences are evolving in the absence of recombination), potential recombinant sequences were detected in the BA.1 and BA.2 subsets using RDP5.23.[40] Specifically each of the ten BA.1 and BA.2 specific datasets were analyzed using default RDP options other than that sequences were considered to be linear. Sequences flagged for signs of recombination with an associated P-value of 0.05 or lower were removed from the datasets. The final datasets for each VOC replicate contained the following numbers of sequences, corresponding to the specified date ranges to match relevant periods of circulation of each variant:

Dataset date ranges (sampling dates):
Alpha: n = 21,280, dates = 2020-10-19 - 2021-07-11
Beta: n = 22,669, dates = 2020-08-19 - 2021-08-07
Gamma: n = 21,331, dates = 2020-09-11 - 2021-09-22
Delta: n = 17,463, dates = 2020-09-07 - 2021-12-23
Omicron BA.1: n = 18,732, dates = 2021-11-17 - 2022-03-08
Omicron BA.2: n = 18,766, dates = 2021-11-27 - 2022-03-15

Phylogenetic reconstruction

For each of the ten replicates per VOC, we produced time scaled tree topologies and performed discrete ancestral state reconstruction (of locations) to infer the global dissemination of each variant. Sequences were aligned using NextAlign (https://github.com/neherlab/nextalign) and Maximum-likelihood tree topologies were inferred using FastTree v.2[41] under a GTR model of nucleotide substitution. The resulting tree topologies were inspected for temporal molecular clock signals using the *clock* functionality of Tree-Time.[42] ML-trees were then transformed into time scaled phylogenies in TreeTime using a standard mutation rate of 0.0008 substitutions per site per year and a standard clock deviation of 0.0004, and using the –confidence flag to get the 90% maximum posterior lower and upper bounds of divergence time estimates and confidence in state transitions in downstream *mugration* analysis. (Using a standard mutation rate resulted in inferences with better confidence compared to using an adjusted mutation rate for each VOC as

determined in a root-to-tip regression analysis (Figures S7D and S7E).) Outlier sequences that deviated from the strict molecular clock assumption as flagged by TreeTime were removed with the Ape package in R[43] until a good time scaled phylogeny was obtained. The *mugration* package extension of TreeTime was then used to map discrete country locations to tips and infer country locations for internal nodes under a GTR model. Finally, a custom python script (available in our GitHub repository: https://github.com/CERI-KRISP/SARS_CoV_2_VOC_dissemination) was used to count the number of state changes over the span of the tree. State transitions that occurred prior to the earliest known tMRCA for each VOC and sub-variant were discarded to minimize the counting of transitions belonging to deep nodes with low confidence

Source sink dynamics and viral movement patterns quantification

From the above ancestral state reconstruction data, phylogeographic maps and source sink dynamics were calculated and plotted using custom R scripts (available in our GitHub repository), as follows. Each recorded location state change is time stamped in decimal dates and annotated with origin and destination countries. Each of those state changes are further annotated with their corresponding continental and sub-continental groupings. Volumes of viral exports and introductions are calculated by aggregating the replicates for each dataset and the mean is considered either per country, continent, sub-continental region and specific to each VOC. Each state change is also annotated to be either a global (between two locations on different continents) or regional viral exchange (between two locations on different continents). Source and sink dynamics are estimated by calculating the net difference between the numbers of exports and introductions for a specific location; a location is determined to be a net source if the number of exports of a variant exceeds the number of introductions. The phylogeographic maps are constructed by linking sub-continental regions with curved lines going anti-clockwise in the direction of the curve. Each curved line is coloured by the mean date of state changes occurring along that specific link.

The speed of VOC arrival in different countries is calculated as the delay (number of days) between the estimated dates of emergence of each VOC from published literature (Table S2) and either the first sampling dates of genomes per VOC and per country as reported on GISAID (Date of access: 18 September 2022) or the first inferred introduction from our ancestral state reconstruction data. GISAID sampling dates are obtained from the curated global metadata file available under the 'Genomic Epidemiology' collection, which is assumed to have undergone minimum sequence quality checks and lineage classification. In addition, we exclude sequences likely to have incorrect sampling dates (where these are either prior to the respective VOC date of emergence specified in Table S2 or where these are over a year prior to the sequences' GISAID submission dates). Pearson correlations are calculated between this delay and mobility into different countries by considering either the total volume of passengers into each country, the volume of passengers from either the UK (for Alpha), South Africa (for Beta and Omicron), Brazil (for Gamma), and India (for Delta) into each country, or an effective distance metric (D_{eff}), for which the calculation is explained below. Since, the phylogenetic methods for estimating the first inferred introduction relies on case-sensitive genomic sampling, therefore on the size of the epidemic, this means that our method will not pick up an early introduction of a variant if this did not rapidly lead to an increase in the number of cases, which we explain in our limitations section. The way to interpret this is that our method is able to estimate the earliest inferred introductions that are relevant to seeding local epidemics. The discrepancy between the dates of first sequenced sample and first inferred introduction is useful to query and interpret (Figure S8). If the first inferred introduction happens after the first sequenced sample, it means that there had already been introductions of the variant much before the introductions that successfully seeded epidemic growth in respective countries. For instance, this is predominantly the case for Delta, where it is known that the variant was spreading before observable epidemic growth (due to still ongoing waves of Alpha, Beta and Gamma locally). If, on the contrary, the first inferred introduction is estimated to be before the first sequenced sample, then it means that the respective country had not yet detected the variant by the time epidemic growth had already been seeded. This is shown to be the case for several of the countries that received the earliest introductions of Alpha, Beta, Gamma and the Omicron sub variants, demonstrating a lag between detection by genomic sequencing and epidemic expansion of variants.

All data visualization was generated through the ggplot package in R.[44]

Air Travel Data

We evaluated travel data generated from the International Air Transport Association (IATA) to quantify passenger volumes originating from international airports during the specified time periods (reported below). IATA data accounts for ~90% of passenger travel itineraries on commercial flights, excluding transportation via unscheduled charter flights (the remainder is modeled using market intelligence). Correlations with air travel passenger volumes were calculated using the Spearman rank correlation method, and reporting levels of significance.

Relevant travel periods:
Alpha: September 2020 - March 2021, Origin: UK
Beta: September 2020 - March 2021, Origin: South Africa
Gamma: November 2020 - May 2021, Origin: Brazil
Delta: September 2020 - September 2021, Origin: India
Omicron BA.1: November 2021 - March 2022, Origin: South Africa
Omicron BA.2: November 2021 - March 2022, Origin: South Africa
Omicon BA.4/BA.5: November 2021 - March 2022, Origin: South Africa

Global travel dataset: February 2020 - March 2022
Complete air travel network: October 2020 & October 2021

Air travel network and sensitivity analysis

We used data from October 2020 and October 2021 when global data on air travel between countries was available. We selected these two months as they represent low travel and 'recovered' passenger travel volumes.

We used global air traffic data from October 2020 and 2021; these two months were selected as they represent broadly the two distinct phases of global mobility patterns: before and after the "recovery" from the substantially reduced intensity of air travel as a result of the pandemic starting in 2019. The data consists of the number of air passengers traveling between 231 countries in the corresponding month.

Calculation of Deff

Following the formulation as detailed in Brockmann and Helbing,[22] from the air passenger matrix F (where Fmn represents the number of air passengers traveling from country m to n during the corresponding month) we first computed the effective length matrix d, where the element dnm is given by

$$d_{mn} \; = \; 1 \; - \; log \; P_{mn}$$

where P_{mn} is the fraction of air passengers leaving country n that are arriving at country m, and therefore can also be written as $P_{mn} = F_{mn}/F_n$, where $F_n = \sum_i F_{in}$ is the total number of air passengers leaving country n.

Having computed the effective length matrix d, we then proceeded to identify the shortest path between the presumed origin location and any other connected node in the network using Dijkstra's algorithm implemented in a Python package. The shortest path between two countries n and m corresponds to the path that traverses a finite set of legs in the network $\tau = \{l_1, l_2, ..., l_L\}$ such that the sum of effective lengths along this set of legs is the smallest among all possible paths from n to m. We define this to be the effective distance D_{mn} from n to m. Note that there is an asymmetry between d_{mn} and d_{nm}, and more importantly $D_{mn} \neq D_{nm}$ in general. For a given presumed original location/node, the tree constructed from the set of shortest paths to all other nodes is known as a shortest path tree.

Comparisons between D_eff and travel volumes

For a given origin location (OL), here we compare the number of air passengers traveling from OL to a country n, $F_{no} = F(OL \rightarrow n)$, versus the corresponding effective distance $D_{no} = D_{eff}(OL \rightarrow n)$. Plotting $log \, F_{no}$ against D_{no} (Figure S6A), we find that all countries either lie on or below a straight diagonal line with a negative gradient. This is unsurprising given the formulation of the effective length where $d_{mn} = 1 - log \, P_{mn}$, which can be rearranged to give $log \, P_{mn} = 1 - d_{mn}$ for any pair of nodes m and n. By the definition of the effective distance D_{mn} which is the sum of the effective lengths along the shortest path from n to m, then $D_{mn} \leq d_{mn}$ and therefore $log P_{mn} \geq 1 - D_{mn}$. Equality is only satisfied when the shortest path from n to m is the direct path between n and m without any intermediate node. With this in mind, the plots of $log \, F_{no}$ against D_{no} therefore show that a large proportion of countries (with observed direct traffic flow from the presumed OL) are connected to OL by their direct path and therefore lie on a straight diagonal line. However, it is important to note that a substantial of countries did not observe direct traffic flow from the presumed OL and are thus not shown in the plots (17% to 40%) - these countries represent peripheral nodes that are not directly connected to OL in the traffic network and are therefore countries that would otherwise have been ignored in an analysis using observed direct traffic flow rather than effective distances.

Comparison between D_eff calculated using data from 2020 and 2021

Comparison between the effective distances calculated using air traffic data from 2020 and 2021 shows that the global mobility patterns were broadly similar across the two years. Spearman's rank correlation coefficients range from 0.86 (with India as the presumed origin location) to 0.93 (with the United Kingdom as the presumed origin location). Visual comparison of the shortest path trees constructed from the shortest paths also reveal mostly similar topological structures across the two years (Figure S5), where most of the nodes retain similar relative positions in the tree and therefore level of importance in the context of global air traffic as a driver of the global dissemination of VOCs.

Sensitivity analysis of United Arab Emirates (ARE) as a major travel hub

We observe from Figure 3B (left) that United Arab Emirates (ARE) potentially played an important role in the global dissemination of the Delta variant, acting as a major travel hub that connects India (IND; the presumed origin location) with many countries in Asia (e.g. Jordan, Pakistan), Africa (e.g. Ethiopia, Egypt) and Europe (e.g. Bosnia-Herzegovina, Serbia). However, we also find that the first sequence of the Delta variant was only detected in ARE in June 2021 - substantially later than other major travel hubs (e.g. GBR, USA) occupying similar positions in the shortest path tree (Figures S6B and S6C), as well as countries that are descendants of ARE (i.e. countries with shortest path from IND that traverses through ARE). This is perhaps not surprising given the very low sequencing intensity in ARE, with only 17 September 2022) since the beginning of the pandemic. To lend further support to the

hypothesis of ARE as an important (yet unobserved phylogenetically) transit node between IND and other countries in the global air traffic network, we performed a sensitivity analysis by recalculating the effective distances and the shortest path tree with ARE removed as an intermediate node. This is equivalent to restricting all outward air traffic from ARE, while allowing inward traffic from other countries into ARE.

From the sensitivity analysis we find that, for the 52 countries that are descendants of ARE in the shortest path tree, there is an average increase in the effective distance from IND by 15.3%, with a range of 1% to 85%. More importantly, we observe substantial structural changes to the shortest path tree upon removal of ARE, where most of these descendant countries no longer form a single cluster but become interspersed across the whole tree, with some smaller clusters forming around other major travel hubs such as GBR and USA. Future work should further investigate the likelihood of the observed position of ARE in the shortest path tree given the observed delays in arrival of the Delta variant as well as other VOCs.

Regression of travel volume, travel restrictions, and case and death counts on viral exports

We employed a negative binomial regression to test the effect of total passenger travel volume, COVID-19 case and deaths counts, and the international travel ban per country on the estimated mean monthly exports of the SARS-CoV-2 virus. The estimated number of viral exports from each country were calculated by aggregating replicates from the ancestral state reconstruction and the mean monthly number of exports calculated per country per VOC. This was used as the response variable in the regression. The predictor variables were as follows: total monthly passenger volume originating from each country of viral export, the total number of reported cases or deaths per month per country (scaled to genomic prevalence of each variant calculated from genomic surveillance data on GISAID), and international travel ban data gathered from the Oxford COVID-19 Government Response Tracker.[47] International travel bans were coded as a categorical variable: 0 - No measures; 1 – Screening; 2 - Quarantine arrivals from high-risk regions; 3 - Ban on high-risk regions; 4 - Total border closure. The strictest travel ban experienced within each calendar month was used as the value for that particular month. Negative binomial regressions were performed, with the MASS R package[48] with a log link function and maximum likelihood estimation of theta, per VOC with data structured at country level and for the period of circulation of each respective VOC. Nagelkerke's pseudo R2 was calculated for each model (DescTools R package[45]) to be cautiously used to assess the proportion of variance for the response variable explained by the predictors.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using R version 4.1.3 and are described in the figure legends and in the method details.
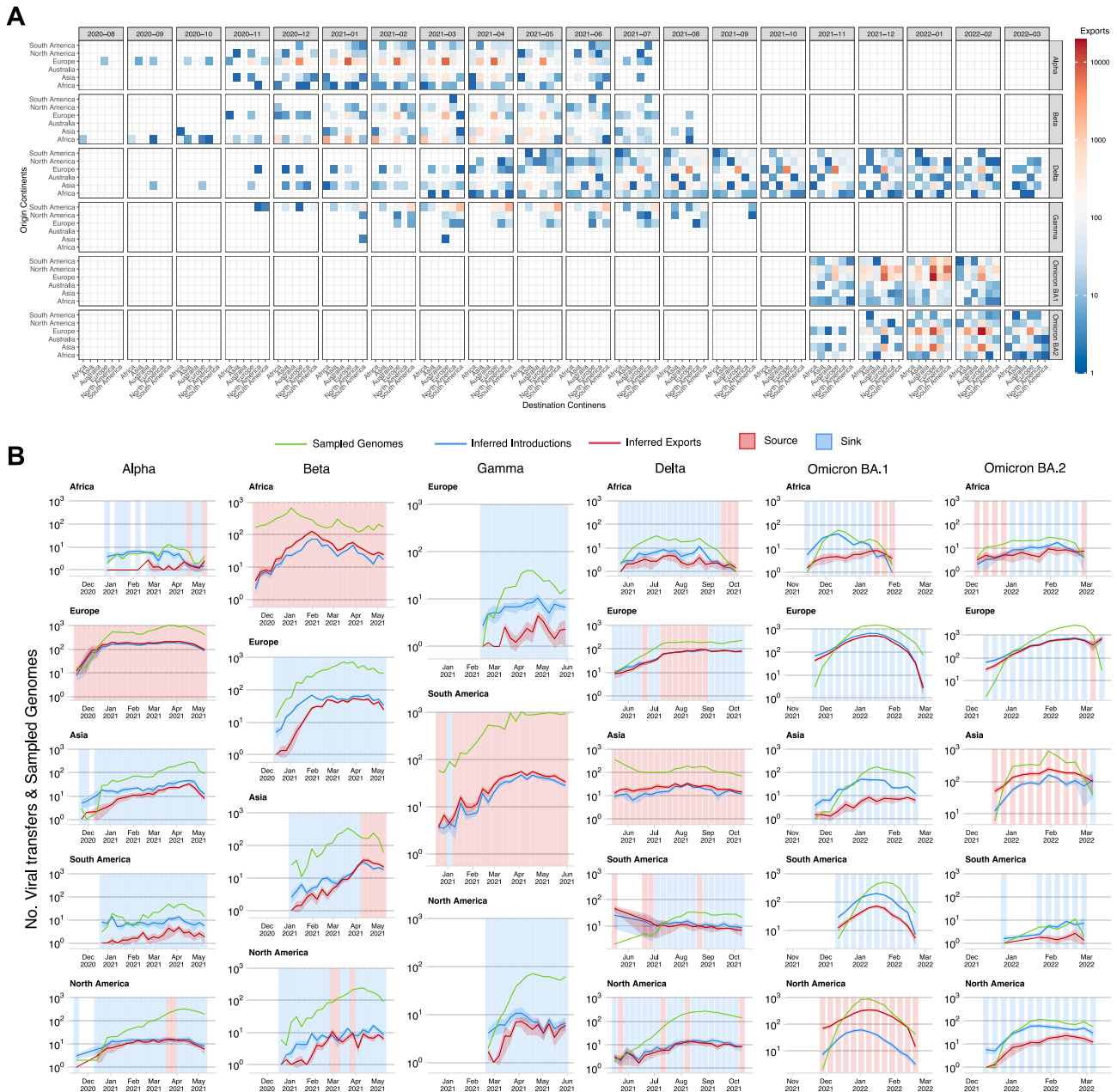
Supplemental figures

Figure S1. Viral sampling, introductions, and exportations for various VOCs over time, related to Figure 1

(A) Connectivity matrix representing the number of VOC-specific exports between continents with a monthly temporal resolution. The number of exports is inferred from case-sensitive phylogeographic analysis (see STAR Methods for details).

(B) Graphs show the time-varying progression of the numbers of sampled genomes in our analysis compared with the numbers of inferred introductions and exportations for Alpha, Beta, Gamma, Delta, Omicron BA.1, and Omicron BA.2 per continent.

A

B

C
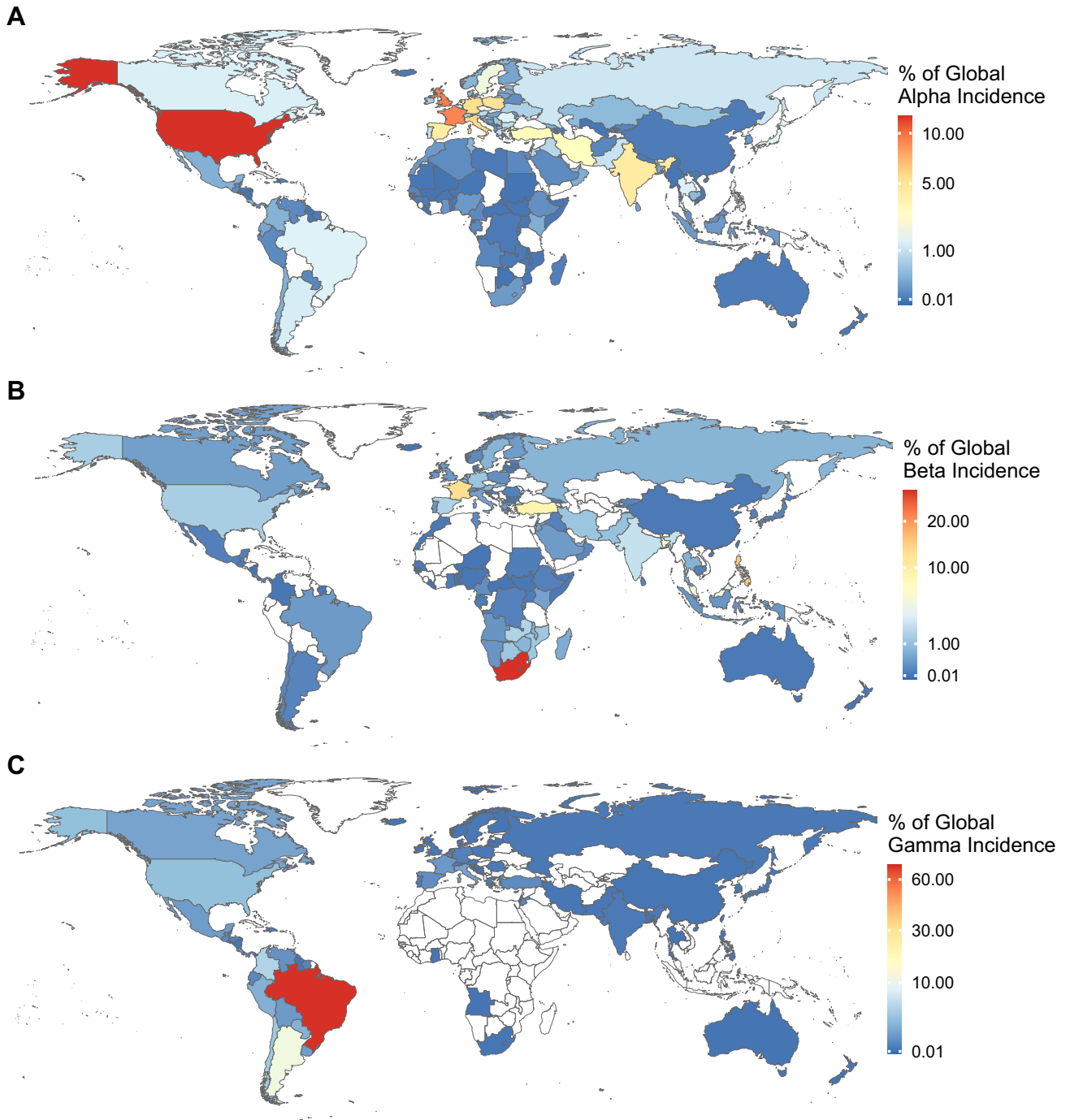
Figure S2. Alpha, Beta, and Gamma global distributions, related to Figure 1
Maps show countries colored by their share of total global Alpha (A), Beta (B), or Gamma (C) incidence.

A

B

C

D
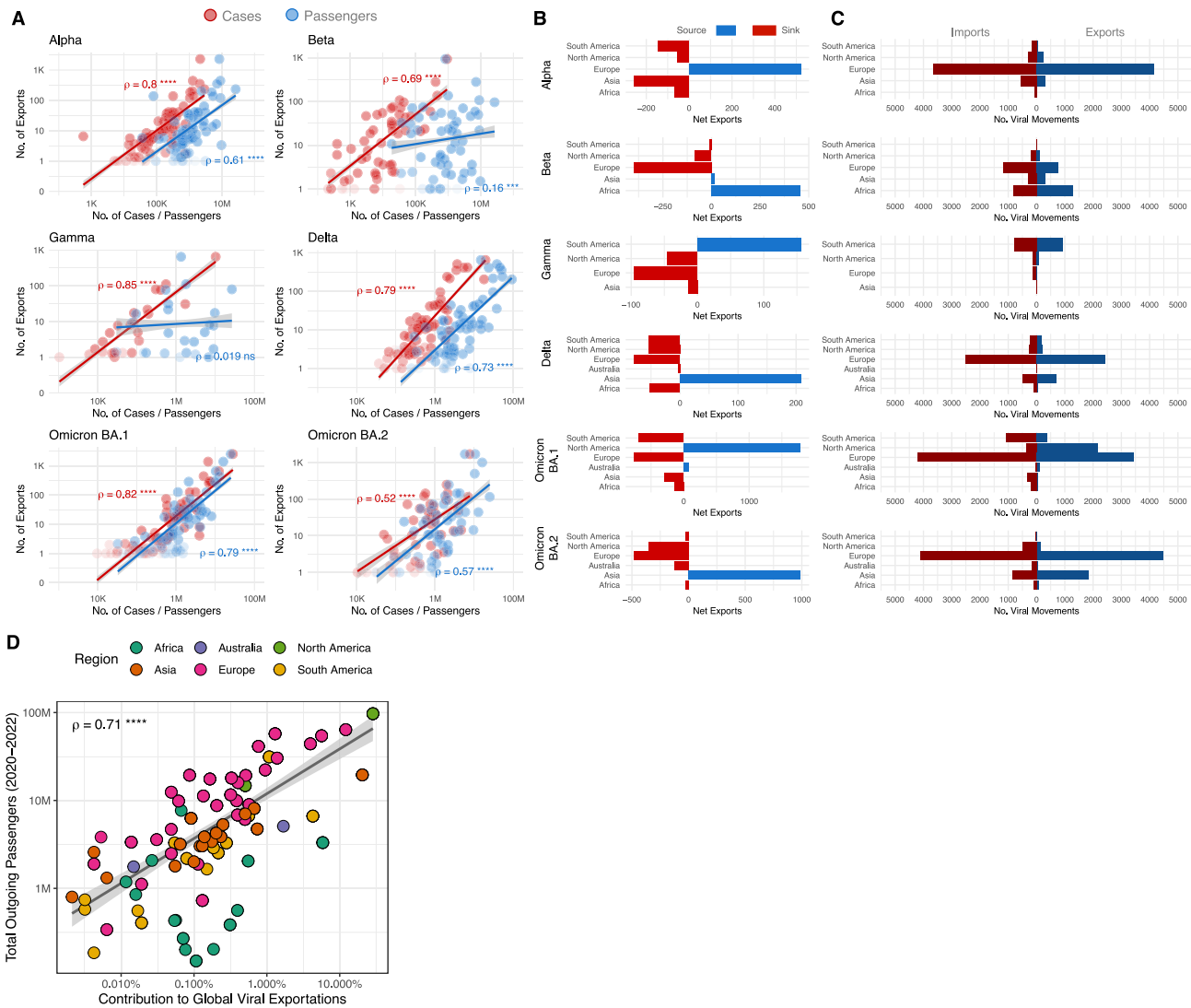
Figure S3. Correlations of incidence and travel to inferred VOC exportation numbers, related to Figures 1 and 2

(A) Graphs show scatter plots and regression lines denoting the numbers of variant-specific cases, volumes of air travel passengers, and inferred numbers of VOC exportations for each country Spearman rank correlation values are shown, with the level of significance indicated.

(B) The net difference between viral exportation and importation events.

(C) The absolute numbers of exportation and importation events for each continent per VOC.

(D) Correlation of contribution of global viral exportation events and outgoing travel from countries. Graph shows a scatter plot and regression line denoting the share of each country's contribution to global numbers of inferred exportations for all VOCs and the total number of outgoing air travel passengers from 2020 to 2022. The Spearman rank correlation value is shown, with the level of significance indicated. The outliers are primarily southern African countries, with global contributions to viral exports comparable to some small European and Asian countries but with visibly lower total outgoing passenger volumes. This can be attributed to the Beta variant, which primarily circulated in southern Africa. The outlier southern African countries shown in this figure likely made significant contributions to Beta exports globally despite their relatively lower connectivity.
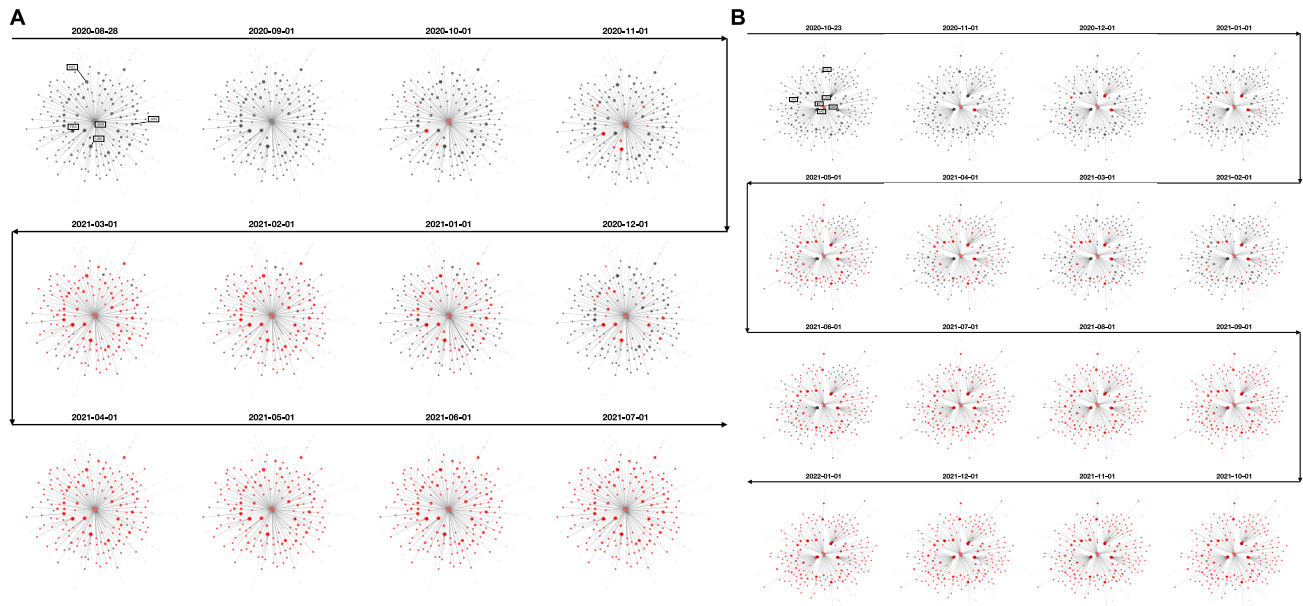
**Figure S4. Global dissemination of the Alpha and Delta variants in effective distance space, related to Figure 4**

(A) The sequence of panels shows the first sampling of the Alpha variant in different countries along the shortest path tree, with the United Kingdom (GBR) as the presumed origin location (OL). Radial distance of each node from the central node represents the effective distance, D_{eff}, from the presumed OL. Each node represents a country and is colored according to whether a sequence of the Alpha variant has been sampled (red) or not (dark gray). Light gray nodes represent countries with either no sampled sequences that are of the Alpha variant or countries that are not connected to the presumed OL in the air traffic network.

(B) The sequence of panels shows the first sampling of the Delta variant in different countries along the shortest path tree, with India (IND) as the presumed origin location (OL). Radial distance of each node from the central node represents the effective distance, D_{eff}, from the presumed OL. Each node represents a country and is colored according to whether a sequence of the Delta variant has been sampled (red) or not (dark gray). Light gray nodes represent countries with either no sampled sequences that are of the Delta variant or countries that are not connected to the presumed OL in the air traffic network.
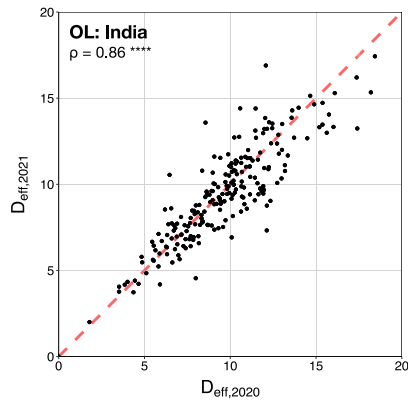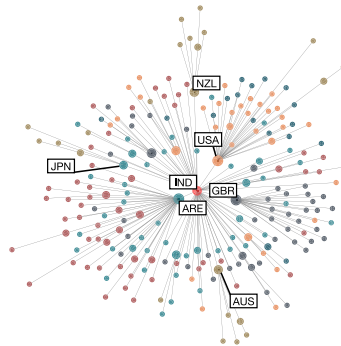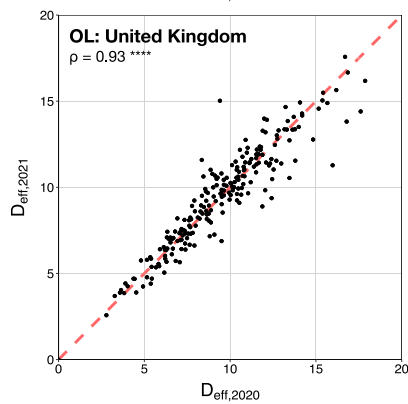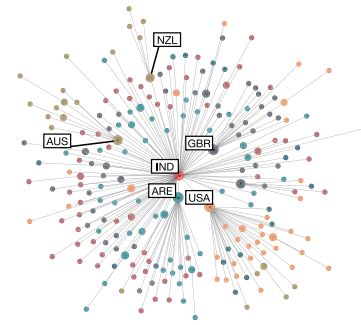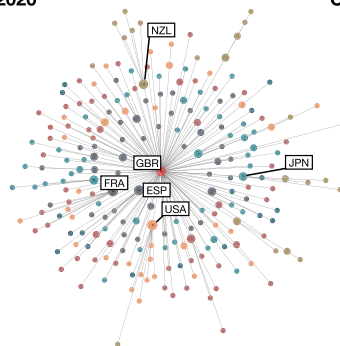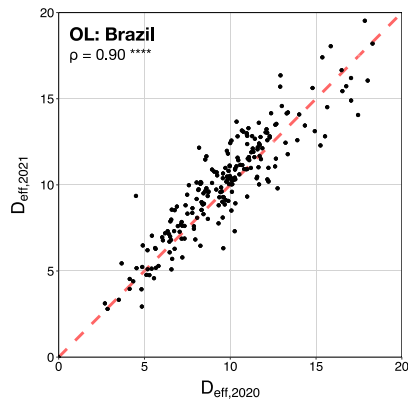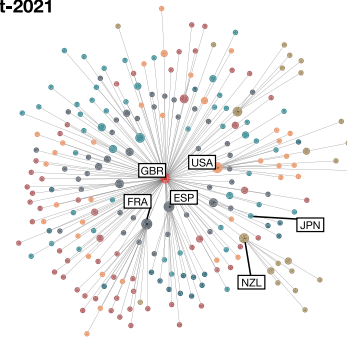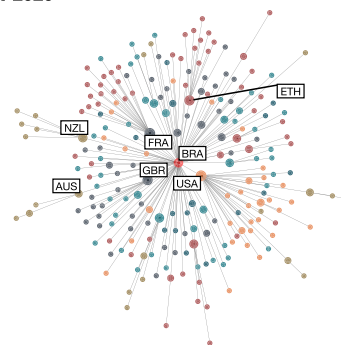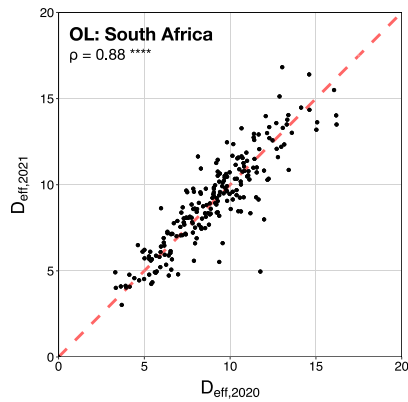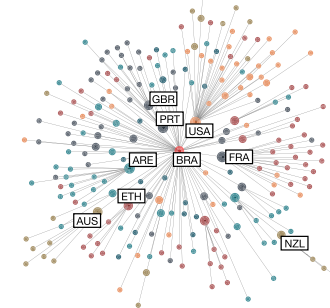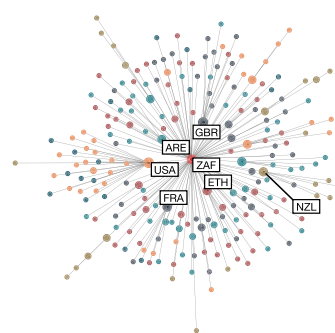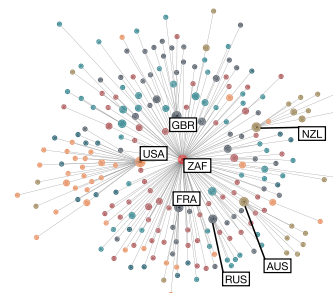
Figure S5. Comparison of global air traffic networks observed in Oct-2020 and Oct-2021, related to Figure 4

Each row of panels corresponds to a presumed origin location associated with one of the VOCs. (Left) Effective distances calculated from global air traffic data observed in Oct-2020 versus Oct-2021. Each black dot represents a country; the red dashed line represents the expected positions of these countries had the air traffic network remained unchanged between 2020 and 2021. (Middle and right) Shortest path trees constructed from the global air traffic observed in Oct-2020 and Oct-2021, respectively. Each circle represents a country, colored according to its corresponding continent. Central red circle represents the presumed origin location. Radial distance of each node from the central node represents the effective distance from the presumed origin location.

A

B

C
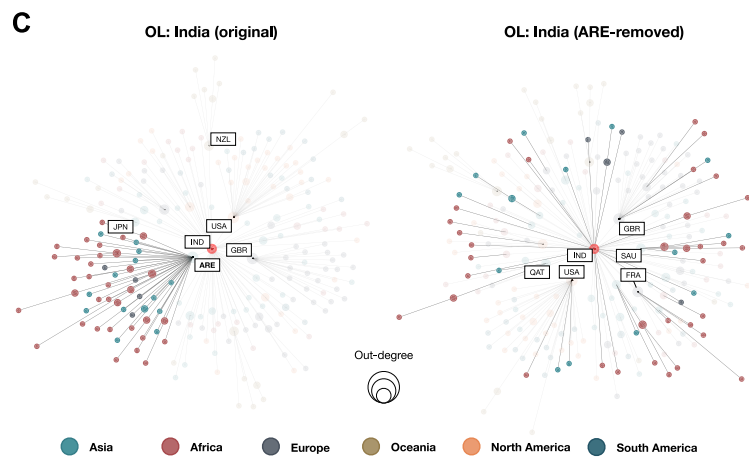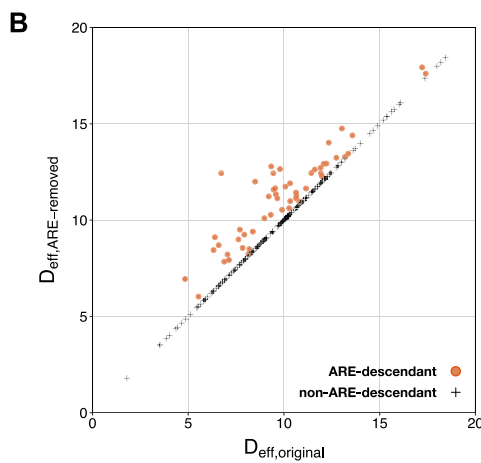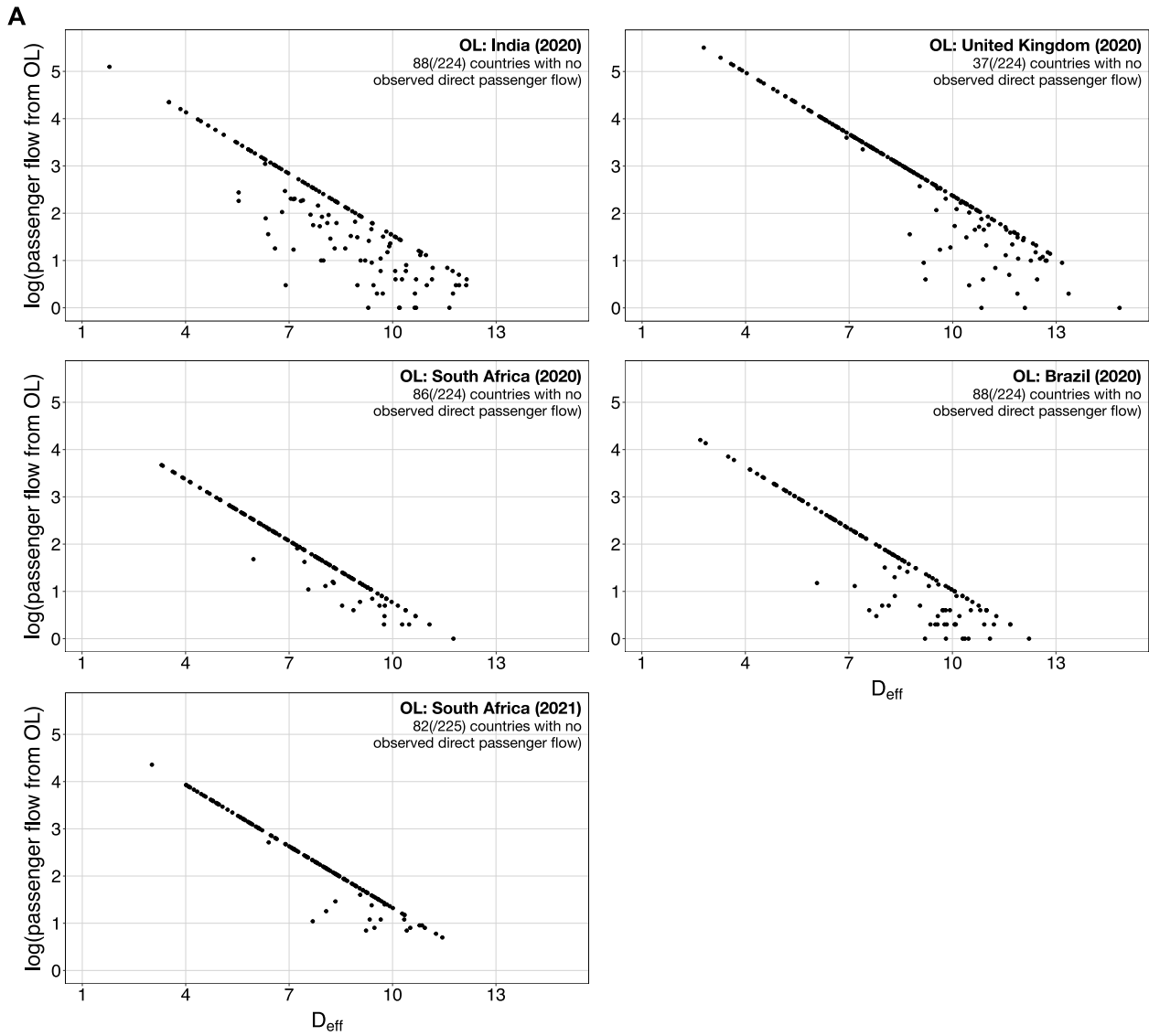
Figure S6. Sensitivity analyses of effective distances, related to Figure 4

(A) Correlation between air passenger flow and effective distances. Log of number of air passengers traveling directly from the presumed origin location (OL) to each country versus effective distance, D_{eff}, calculated from the global air traffic network relevant to each VOC. Number of countries with no observed direct passenger flow from the presumed OL is indicated in each panel.

(B and C) Sensitivity analysis of the United Arab Emirates as a secondary transit hub in the global dissemination of the Delta variant.

(B) Effective distance before versus after the United Arab Emirates (ARE) is removed as an intermediate node in the air traffic network. Orange circles represent countries that are descendants of ARE in the shortest path tree, i.e., countries with shortest path that traverses from India through ARE. Black crosses represent countries that are not descendants of ARE and therefore have an effective distance that is unaffected by the removal of ARE.

(C) Shortest path tree before (left) and after (right) the removal of ARE as an intermediate node. Highlighted nodes represent countries that are descendants of ARE prior to the removal. Red node at the center represents India, the presumed origin location of the Delta variant.
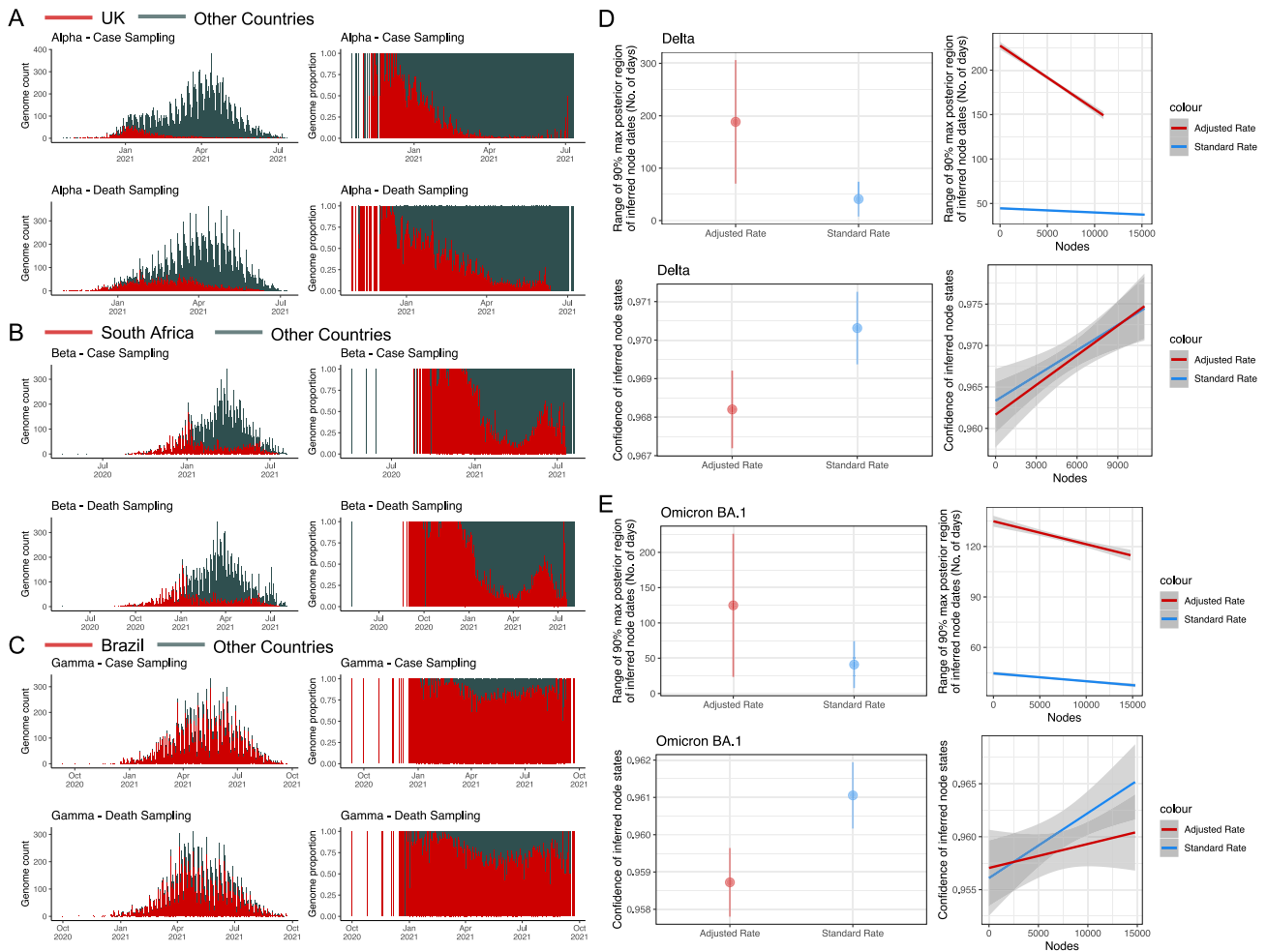
Figure S7. Methodology sensitivity analysis, related to STAR Methods

(A–C) Genomic counts and proportions, or presumed origin country vs. other countries for Alpha, Beta, and Gamma, when genomic subsampling is performed either proportional to VOC-specific case counts or VOC-specific deaths. This comparison was performed to ascertain potential biases in testing rates in the resulting sampling proportions. We found that the sampling proportions for each country remains consistent whichever strategy is used and opt to perform the rest of the analysis with case sampling.

(D and E) Justification for the use of evolutionary rates. Graphs show the range of 90% maximum posterior region of inferred node dates (in number of days) and the confidence of reconstructed node states as proxies for robustness of inference, either as an averaged measure for all nodes or by node number, from deepest nodes for the adjusted evolutionary rate vs. the standard evolutionary rate. Results are shown for one phylogenetic reconstruction of Delta and Omicron BA.1 datasets.
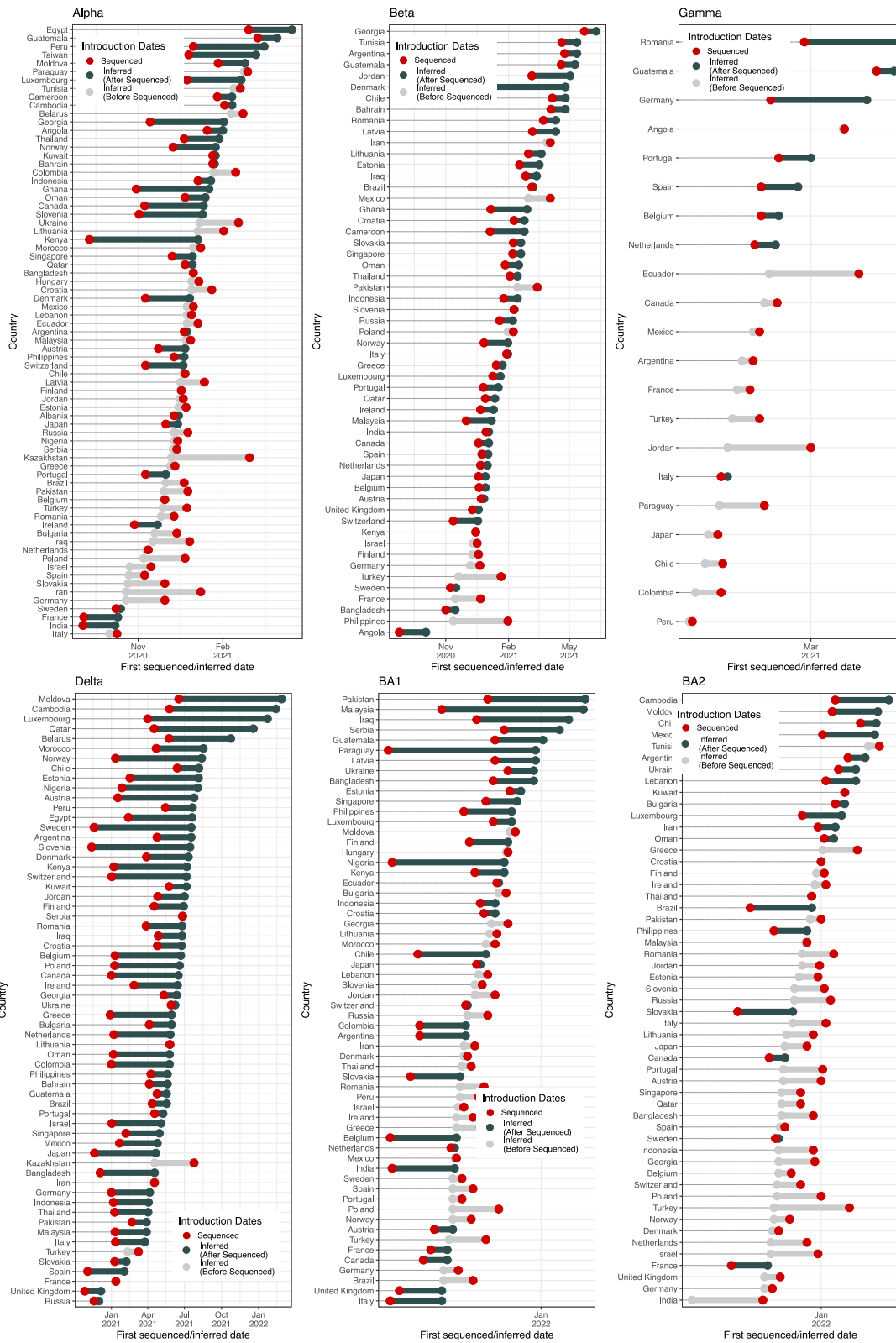
(legend on next page)

Figure S8. VOC introduction dates per country, shown either as the first sequenced date on GISAID or the date of first inferred introduction from our phylogenetic analysis, related to discussion

The first sequenced dates are shown as a red circle, and the first inferred introductions are shown either in dark gray if they happen after the first sequenced date or light gray if it happens before. If the first inferred introduction happens after the first sequenced sample, it means that there had already been introductions of the variant much before the introductions that successfully seeded epidemic growth in respective countries. For instance, this is predominantly the case for Delta, where it is known that the variant was spreading before observable epidemic growth (due to still ongoing waves of Alpha, Beta, and Gamma locally). If, on the contrary, the first inferred introduction is estimated to be before the first sequenced sample, then it means that the respective country had not yet detected the variant by the time epidemic growth had already been seeded. This is shown to be the case for several of the countries that received the earliest introductions of Alpha, Beta, Gamma, and the Omicron subvariants, demonstrating a lag between detection by genomic sequencing and epidemic expansion of variants.