

# Comparative phylodynamics reveals the evolutionary history of SARS-CoV-2 emerging variants in the Arabian Peninsula

Moh A. Alkhamis,<sup>1,\*</sup>†,‡ Nicholas M. Fountain-Jones,<sup>2,†,§</sup> Mohammad M. Khajjah,<sup>3</sup> Mohammad Alhounaim,<sup>4,5</sup> and Salman K. Al-Sabah<sup>5,6,\*</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Faculty of Public Health, Health Sciences Centre, Kuwait University, Street 109, Jabriya Campus, Safat, Kuwait City 13110, Kuwait, <sup>2</sup>School of Natural Sciences, University of Tasmania, Hobart, Dynnyrne, TAS 7000, Australia, <sup>3</sup>Systems and Software Development Department, Kuwait Institute for Scientific Research, Al-Jaheth Street, Shuwaikh, Safat, Kuwait City 13109, Kuwait, <sup>4</sup>Departement of Pediatrics, Amiri Hospital, Ministry of Health, Main street, Kuwait City 13041, Kuwait, <sup>5</sup>Jaber Al-Ahmad Al-Sabah Hospital, Ministry of Health, South Surra street, Kuwait City 13094, Kuwait and <sup>6</sup>Department of Surgery, Faculty of Medicine, Health Sciences Center, Kuwait University, Street 109, Jabriya Campus, Safat, Kuwait City 13110, Kuwait

†Authors contributed equally to this work.

\*<https://orcid.org/0000-0001-6451-5247>

§<https://orcid.org/0000-0001-9248-8493>

‡<https://orcid.org/0000-0002-8081-8544>

\*Corresponding author: E-mail: [m.alkhamis@ku.edu.kw](mailto:m.alkhamis@ku.edu.kw)

## Abstract

Emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants continue to be responsible for an unprecedented worldwide public health and economic catastrophe. Accurate understanding and comparison of global and regional evolutionary epidemiology of novel SARS-CoV-2 variants are critical to guide current and future interventions. Here, we utilized a Bayesian phylodynamic pipeline to trace and compare the evolutionary dynamics, spatiotemporal origins, and spread of five variants (Alpha, Beta, Delta, Kappa, and Eta) across the Arabian Peninsula. We found variant-specific signatures of evolution and spread that are likely linked to air travel and disease control interventions in the region. Alpha, Beta, and Delta variants went through sequential periods of growth and decline, whereas we inferred inconclusive population growth patterns for the Kappa and Eta variants due to their sporadic introductions in the region. Non-pharmaceutical interventions imposed between mid-2020 and early 2021 likely played a role in reducing the epidemic progression of the Beta and the Alpha variants. In comparison, the combination of the non-pharmaceutical interventions and the rapid rollout of vaccination might have shaped Delta variant dynamics. We found that the Alpha and Beta variants were frequently introduced into the Arab peninsula between mid-2020 and early 2021 from Europe and Africa, respectively, whereas the Delta variant was frequently introduced between early 2021 and mid-2021 from East Asia. For these three variants, we also revealed significant and intense dispersal routes between the Arab region and Africa, Europe, Asia, and Oceania. In contrast, the restricted spread and stable effective population size of the Kappa and the Eta variants suggest that they no longer need to be targeted in genomic surveillance activities in the region. In contrast, the evolutionary characteristics of the Alpha, Beta, and Delta variants confirm the dominance of these variants in the recent outbreaks. Our study highlights the urgent need to establish regional molecular surveillance programs to ensure effective decision making related to the allocation of intervention activities targeted toward the most relevant variants.

**Key words:** SARS-CoV-2; variants of concern; phylodynamics; phylogeography; genomic surveillance.

## 1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused unprecedented health and economic losses across the globe (Andersen et al. 2020). Within approximately 4 months from SARS-CoV-2 emergence in China in late 2019 (Zhu et al. 2020), it rapidly spread to almost every country, causing unprecedented rapid hospitalizations and mortalities on global scale [World Health Organization (WHO) 2021a]. Thus, on 11 March 2020, the World Health Organization (WHO) declared the new

disease a pandemic (WHO Emergency Committee. Statement on the second meeting of the International Health Regulations 2005), which subsequently prompted governments to impose unparalleled public health restrictions that lasted for approximately 2 years. As of October 2021, over 240 million cases and 5 million deaths were reported in 223 countries or territories [World Health Organization (WHO) 2021a]. The SARS-CoV-2 is an enveloped single-stranded positive-sense RNA virus that causes the coronavirus disease 2019 (COVID-19) in humans (genus *Betacoronavirus*,

family *Coronaviridae*). This family of viruses infects a wide variety of mammalian and avian species, and SARS-CoV-2 is the seventh coronavirus recognized to infect humans from potentially zoonotic origins (Andersen et al. 2020; Zhu et al. 2020). As of December 2021, RaTG13 and RmYN02 coronaviruses are the closest relatives to the SARS-CoV-2 (96 per cent and 93 per cent nucleotide similarity, respectively) and are commonly isolated from wild bats belonging to the species *Rhinolophus affinis* and *Rhinolophus malayanus* (Zhou et al. 2020).

The Gulf Cooperation Council (GCC) countries across the Arabian Peninsula have accumulated a total of 2.5 million confirmed cases and 19.5 thousand fatalities as of 20 October 2021 [World Health Organization (WHO) 2021a]. The first confirmed four cases of SARS-CoV-2 in the GCC region were detected in the United Arab Emirates on 29 January 2020, in a family of travelers arriving from Wuhan in central China (Nandkeolyar 2020). Subsequently, Kuwait, Bahrain, Qatar, Oman, and Saudi Arabia reported their first confirmed cases between mid-February and early March of 2020 [World Health Organization (WHO) 2021a]. Notably, patterns of SARS-CoV-2 prevalence are heterogeneous among GCC member states. For example, between 29 January 2020 and 20 October 2021, the highest prevalence of SARS-CoV-2 cases was reported in Bahrain (16.24 per cent), whereas the lowest was reported in Saudi Arabia (0.02 per cent) [World Health Organization (WHO) 2021a]. Yet, such observed prevalence was highly correlated with country population size, number of daily tests, and whether the countries reported both symptomatic and asymptomatic cases or only symptomatic patients. Furthermore, the mortality rate from the disease during the period, described above, was low (< 0.001 per cent) in all GCC member states compared to the European countries such as Spain and Italy [World Health Organization (WHO) 2021a]. This low mortality rate from SARS-CoV-2 is mainly attributed to the well-funded healthcare system (Al Awaidy et al. 2021) and the region's past experience with the successful control and prevention efforts against the Middle East Respiratory Syndrome (Assiri et al. 2013).

Like many countries, the GCC member states imposed a series of strict non-pharmaceutical interventions throughout the pandemic period, including complete and partial lockdowns, international travel restrictions, public gatherings, contact tracing, and school closures (Hale et al. 2021). During the first wave, these non-pharmaceutical interventions were considered successful in reducing the number of cases, hospitalizations, and deaths [World Health Organization (WHO) 2021a]. However, for the subsequent waves, which were dominated by emerging SARS-CoV-2 variants, non-pharmaceutical interventions were less effective in shaping the epidemic curves in the GCC member states. Declines in cases, hospitalization, and mortalities only occurred when most GCC countries reached a vaccination coverage of greater than 40 per cent [World Health Organization (WHO) 2021a].

After the first global wave of the pandemic, new variants of the SARS-CoV-2, characterized by increased transmissibility and a large number of mutation and deletions rates on the spike (S) protein, began to rapidly emerge and spread within a short time in different continents (Gomez, Perdiguero, and Esteban 2021). These variants are classified differently by the major repositories and organizations; Global Initiative on Sharing All Influenza Data (GISAID; gisaid.org), NextStrain (nextstrain.org), and Pangolin (cov-lineages.org). However, the WHO reclassified SARS-CoV-2 emerging variants into three categories, namely variants of concern (VOC), variant of interest (VOI), and variants under monitoring (VUM) [World Health Organization (WHO) 2021b]. As of October 2021, VOCs include Alpha (B.1.1.7, GRY, or

20I), Beta (B.1.351, GH/501Y.V2, or 20H), Gamma (P.1, GR/501YV3, or 20J), and Delta (B.1.617.2, or 21A, 21I, and 21J), whereas VOIs includes Lambda (C.37, GR/452Q.V1, or 21G) and Mu. Additionally, VUMs includes Kappa (B.1.617.1, G/452R/V3, or 21B), Iota (B.1.526, GH/253.V1), Eta (B.1.525, G/484K.V3, or 20A/S484K), and Epsilon (B.1.427/B.1.429, GH/452R/V1, or 20C/S.452R) [World Health Organization (WHO) 2021b]. The WHO issued guidance for each designated variant category containing recommended surveillance activities and primary action measures [World Health Organization (WHO) 2021b]. We adapted the nomenclature proposed by the WHO throughout the study to maintain its clarity and consistency.

Many countries have implemented SARS-CoV-2 whole-genome sequencing programs as a genetic surveillance tool throughout the pandemic, including the GCC member states (Tayoun et al. 2020; Al-Mahruqi et al. 2021; Benslimane et al. 2021; Bindayna, Deifalla, and Mokbel 2021; Liu et al. 2021; Mourier et al. 2022). However, countries that implemented integrated phylodynamic methods have uncovered important epidemiological insights at local, regional, and global scales (Benvenuto et al. 2020; Fountain-Jones et al. 2020; Worobey et al. 2020). For example, Bayesian phylodynamic methods have been used as a robust tool for investigating circulating variants' evolutionary epidemiology by tracing their origins (Giovannetti et al. 2020; Nabil, Sabrina, and Abdelhakim 2020; Butera et al. 2021), quantifying transmission events (Miller et al. 2020; Seemann et al. 2020), and assessing the impact of non-pharmaceutical interventions before and after their implementation (Geoghegan et al. 2020; López et al. 2021). However, phylodynamic and biogeographic methods have not been fully utilized in the GCC member states, with the exception of Saudi Arabia (Mourier et al. 2022), as well as in other Asian countries (Farah et al. 2020; Jeewandara et al. 2021). This represents an important gap not only for local COVID-19 surveillance but also as the region is a major travel hub between Europe and Asia (Tayoun et al. 2020). Between 16 November 2020 and 28 June 2021, Alpha, Beta, Delta, Kappa, and Eta were the most prevalent WHO-designated variants in the GCC member states, but the epidemiological dynamics of these variants is largely unknown in the region. Utilized in a unified Bayesian framework (Pybus and Rambaut 2009; Grubaugh et al. 2019), these methods have the capacity to model viral evolution and spread in time and space to untangle epidemiological patterns and variant dynamics.

Between November and December 2021, the Omicron variant (B.1.1.529, or GRA) emerged and rapidly spread to over 70 countries worldwide [World Health Organization (WHO) 2021b]. Thus, a comparative understanding of global and regional evolutionary epidemiology of SARS-CoV-2 rapidly emerging variants is critical to guide current and future interventions. Here, we adapt a Bayesian phylodynamic pipeline to trace and compare evolutionary history, spatiotemporal origins, and transmission dynamics of the five variants described above. We uncover new comparative insights into the evolutionary characteristics of SARS-CoV-2 selected variants and their spread from and into the GCC member states. These findings may improve current genetic surveillance efforts on regional scales and subsequently reduce the impacts of current and future emerging variants.

## 2. Methods

### 2.1 Preliminary phylogenetic analysis

We retrieved 3728 non-identical complete genome SARS-CoV-2 sequences from the GISAID (file S1) isolated from cases reported in the GCC member states between 28 January 2020 and 28

June 2021. We excluded identical sequences from the analysis, as they usually have a very low prevalence in full genome datasets. Also, 100 per cent identical sequences usually do not improve the temporal signal of the phylogeny, lead to improper convergence of the posterior estimates, and may substantially increase the computational costs (Lycett et al. 2021; Morel et al. 2021). Approximately 81.5 per cent ( $n=3041$ ; file S1) of the GCC sequences contained earlier lineages such as A, B, C, and G. While approximately 18.5 per cent (686; Supplementary Table S1) of the published sequences contained WHO-designated variants, which comprised the Alpha, Beta, Delta, Kappa, and Eta. Next, we downloaded the metadata of the sequences used to construct the Novel Coronavirus-Global Subsampling (n-cov) tree from the NextStrain web interface (<https://nextstrain.org/ncov/gisaid/global>) in order to retrieve the corresponding sequences from the GISAID database (file S2 for the accession numbers). The NextStrain sequence dataset comprised 2971 globally representative sequences collected between 26 December 2019 and 1 July 2021. Here, we adapted the terminology suggested by NextStrain documentation (Hadfield et al. 2018) and designated the GCC dataset as focal sequences (i.e. sequences obtained locally), while the NextStrain global dataset as context sequences (i.e. sequences obtained from global public repository). The NextStrain algorithm randomly subsamples approximately 600 sequence predefined continental regions, with approximately 400 from the previous 4 months and approximately 200 from earlier strains, to generate representative inferences that are computationally efficient for the average users (see <https://nextstrain.org/ncov/gisaid/global>). Since countries like Denmark have substantially surpassed other regions in terms of the number of sequences isolates (see [gisaid.org](https://gisaid.org)). We chose the NextStrain genomic dataset as our context sequences because they ensure a more equitable representation of sequence distribution across different continental regions.

To reduce the computational intensity of the subsequent analyses, we used genome-sampler version 2.0 (Bolyen et al. 2020) to subsample the context sequences while maintaining their representation in terms of the time of isolation, location of isolation, and genetic diversity (Bolyen et al. 2020). This procedure excluded 1781 sequences (59.9 per cent) of the context sequences, and therefore, the final combined dataset (i.e. focal and context) comprised 4918 sequences. We used MAFFT version 7.49 (Katoh and Standley 2013) to align our sequence dataset and trimmed the first 130bp and last 50bp from the final alignment, as suggested elsewhere, using AliView version 1.74 (Larsson 2014). We found no recombination events in the final alignment using Recombination Detection Program version 4.0 (Martin et al. 2015). We then used IQ-tree version 2.0 (Minh et al. 2020) to construct a maximum likelihood (ML) tree with 1000 ultrafast bootstraps. Using the inbuilt model selection algorithm implemented in IQ-tree (Kalyaanamoorthy et al. 2017), we found the general time-reversible model with empirical base frequencies and gamma-distributed varying rates with four rate categories (GTR + F + R4) was the best substitution model fitting the data.

We then extracted the sequences from each WHO-designated variant (i.e. Alpha, Beta, Delta, Kappa, and Eta; Fig. 1) from the ML tree and constructed an independent ML tree for each variant. We used the root-to-tip regression procedure implemented in TempEst version 1.5.3 (Rambaut et al. 2016) to confirm the presence of significant temporal in the phylogeny of each variant. We found that all variants' sequence datasets exhibited significant temporal signals ( $R^2 > 0.2$ ; Fig. 1), suitable for the subsequent molecular

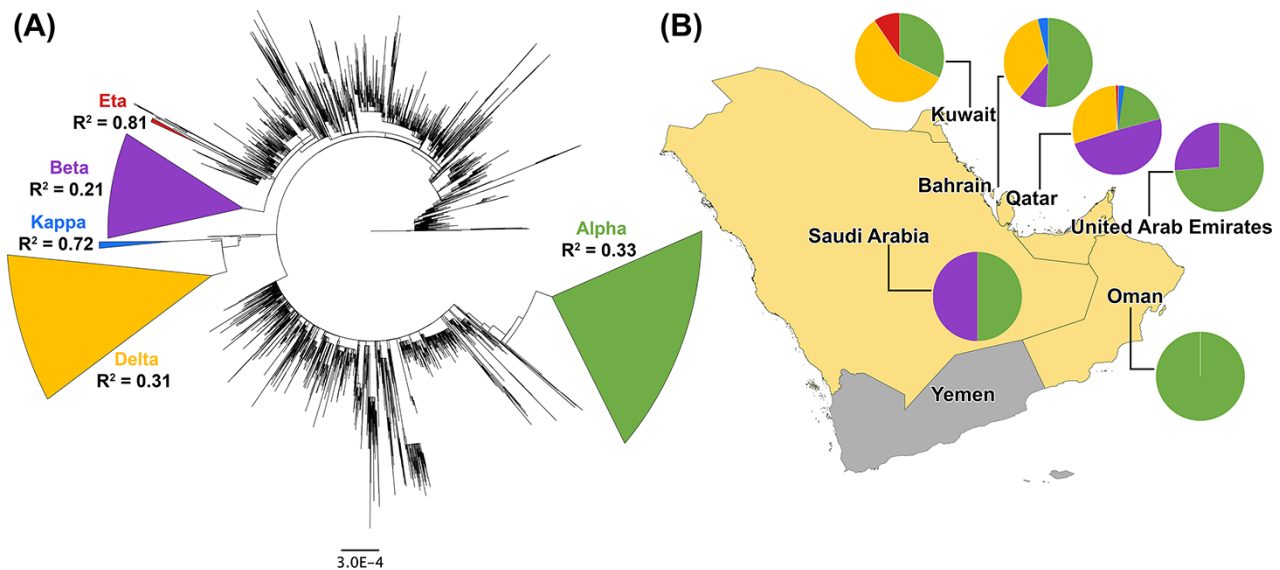
clock analyses. Furthermore, we also used TempEst to investigate additional sequence errors and outliers within the dataset of each variant. Additionally, we used TreeTime (Sagulenko, Puller, and Neher 2018) to calibrate the phylogeny (not shown) of each variant in time. We used two genomes from Wuhan 2019 (i.e. hCoV-19/Wuhan/WH01/2019 and hCoV-19/Wuhan/Hu-1/2019) as an outgroup to estimate the posterior parameters of the subsequent molecular clock analyses.

## 2.2 Variant demographics and phylogeography

We reconstructed the demographic and phylogeographic history of SARS-CoV-2 selected variants using the relaxed-molecular clock models implemented in BEAST 1.10.4 (Suchard et al. 2018) and the BEAGLE library (Ayres et al. 2019) to improve the computational efficiency of the subsequent phylodynamic models. We used the substitution model described above and assessed three parametric [i.e. constant population size (Kingman 1982), exponential (Griffiths and Tavaré 1994), and expansion growth (Griffiths and Tavaré 1994)] one non-parametric [i.e. Bayesian Skygrid (Gill et al. 2013)] coalescent node-age tree priors. Also, we evaluated the uncorrelated exponential and the lognormal branch-rate priors in combination with each node-age model. Additionally, we compared eight candidate models using the Alpha and Delta sequence datasets to determine the best fitting demographic model for other variants in the subsequent analyses, using the Bayes factor (BF) comparison of the marginal-likelihood estimates calculated using path-sampling (Ps) and stepping-stone (Ss) methods (Baele et al. 2012).

We quantified the global origins of SARS-CoV-2 selected variants and identified significant dispersal routes between the GCC and other geographical regions (i.e. continents) using the discrete Bayesian phylogeographic analysis extension implemented in BEAST (Lemey et al. 2009). We aggregated the GCC countries into a single geographical location and specified a total of seven discrete traits, including Africa, Asia, Europe, GCC, Oceania, and North and South America (Supplementary Table S2). However, for each variant, continents with sequences less than two were excluded from the discrete trait assignment to limit the impact of biased posterior inferences related to the phylogeographic analysis (De Maio et al. 2015). For each variant, we identified significant dispersal routes and their directionality between regions using the Bayesian stochastic search variable selection (BSSVS) approach (Pybus and Rambaut 2009). Also, we explored the fit of each sequence dataset to the asymmetric model (irreversible transitions) and the symmetric model (reversible transitions) using the BF comparisons, as described above. Further, we used the Markov-jump (MJ) procedure (Farr, Mandel, and Stevens 2015) to infer the expected number of backward and forward transitions between the GCC and other regions.

We performed our phylodynamic analysis in duplicate and ran the Markov chain Monte Carlo (MCMC) simulations for 350 million iterations sampling every 30,000 steps for the Alpha, Beta, and Delta variants. However, due to the small sample sizes of the Kappa and Eta datasets, we performed 100 million MCMC cycles while sampling every 10,000 steps. We used Tracer version 1.7.1 (Rambaut et al. 2018) to monitor the proper convergence of the posterior parameters by reaching the effective sample size (i.e. ESS > 200) We discarded the first 10 per cent of the sampled trees from each MCMC chain and generated the maximum clade credibility (MCC) tree based on the median node heights using TreeAnnotator version 1.10.4.



**Figure 1.** Reconstructed ML ( $n = 4918$ ) tree and the selected variants' prevalence among the Arabian Peninsula countries of SARS-CoV-2 between November 2020, and June 2021. (A) The ML tree consisted of all SARS-CoV-2 variants identified worldwide between 26 December 2019, and 28 June 2021 and was reconstructed using the GTR + F + R4 substitution model, and 1000 bootstrap replicates implemented in IQ-Tree version 2. The scale bar below the tree indicates the substitution rate per site. The five selected variants are highlighted by distinct colors, and their root-to-tip divergence was estimated using TempEst version 1.5.3. (B) Pie charts represent the prevalence of the selected variants per GCC country, in which the size of the pies is proportional to the number of sequences published in [www.gisaid.org](http://www.gisaid.org). The maps were generated using Q-GIS version 3.16.

To estimate the effective population size of each variant through time, we used the Skygrid coalescent tree with the uncorrelated lognormal branch-rate models. We used FigTree version 1.4.4 (Rambaut 2018) to plot the resulting phylogeographic MCC tree for each variant. We estimated the effective population size for all sequences combined as well as for clades circulating predominantly in the GCC. We identified these clades based on the root-state posterior probability inferred from the phylogeographic reconstructions that represent the earliest introduction into the GCC region. Clades introduced to the GCC that had >5 per cent of the sequences from other countries were excluded. We did not use this approach on the Kappa and the Eta variant due to the extremely low number of sequences isolated in the GCC (Supplementary Table S1). Then, we used Tracer to generate the Skygrid plots for each variant as well as for each identified dominant GCC clade. Further, we used SPREAD3 (Bielejec et al. 2016) version 0.9.6 to assess the significance of the non-zero rate dispersal routes using the implemented BSSVS BF procedure (BSSVS BFs > 3). We plotted significant dispersal routes between regions using the JavaScript Object Notation maps and SPREAD3. Finally, we used the Bayesian Tip-association significance test (Parker, Rambaut, and Pybus 2008) implemented in BaTs version 2.0 to validate the association between the posterior phylogeny and spatial discrete traits. We used the association index (AI) and parsimony score (PS) computed from a sample of 1000 posterior trees and a 100 null replicated to assess whether the sampling regions contributed significantly to the structure of the inferred MCC tree for each variant.

### 3. Results

#### 3.1 Demographic history and pandemic dynamics

The results of the BF comparisons indicated that the combination of the expansion growth node-age and the uncorrelated exponential branch-rate priors are the best fitting model for the

**Table 1.** Inferred substitution per site per year for the selected SARS-CoV-2 variants in the GCC countries between 16 November 2020 and 28 June 2021.

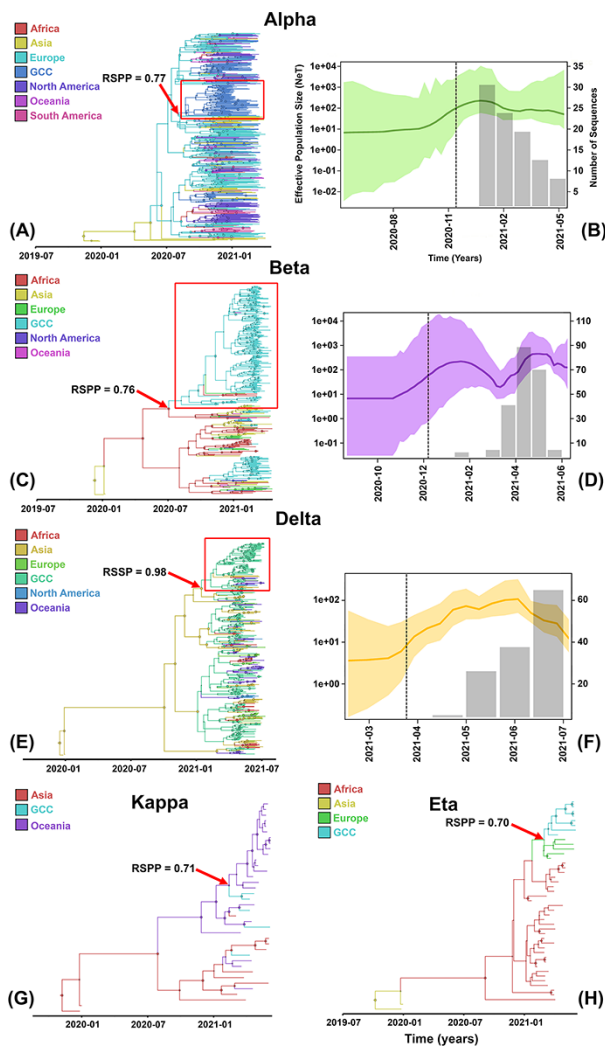
Variant	Substitution/site/year	95% HPD <sup>a</sup>
Alpha	$2.60 \times 10^{-4}$	$[1.63 \times 10^{-4}, 3.70 \times 10^{-4}]$
Beta	$7.0 \times 10^{-4}$	$[5.32 \times 10^{-4}, 8.68 \times 10^{-4}]$
Delta	$1.12 \times 10^{-3}$	$[9.71 \times 10^{-4}, 1.38 \times 10^{-3}]$
Eta	$5.07 \times 10^{-4}$	$[4.81 \times 10^{-4}, 8.78 \times 10^{-4}]$
Kappa	$6.49 \times 10^{-4}$	$[4.63 \times 10^{-4}, 9.09 \times 10^{-4}]$

<sup>a</sup>HPD = High Posterior Density.

Alpha and Delta variants (BFs > 50 and 3, respectively; Tables S3 and S4) and therefore was used to infer the evolutionary history of other variants. The highest estimate of the mean nucleotide substitution rate per year was inferred for the Delta variant ( $1.12 \times 10^{-3}$ , 95 per cent HPD  $[9.71 \times 10^{-4}, 1.38 \times 10^{-3}]$ ; Table 1), whereas the lowest estimate was inferred for the Alpha variant ( $2.60 \times 10^{-4}$ , 95 per cent HPD  $[1.63 \times 10^{-4}, 3.70 \times 10^{-4}]$ ; Table 1). Our demographic reconstructions revealed sequential peaks of the effective population size for Alpha, Beta, and Delta variants (Fig. 2).

Our BF comparisons revealed that the asymmetric phylogeographic model with irreversible transitions was the best fitting model for the Alpha and the Delta variants. In contrast, the symmetric phylogeographic model with reversible transitions was the best fitting model for the Beta, Kappa, and the Eta variants (BFs > 15). Further, all of the estimated AI indices and PS scores of the selected geographic discrete traits were statistically significant among all variants (Table 2).

Our combined phylogeographic and demographic reconstructions revealed sequential variant waves with divergent phylogeographic and demographic histories (Fig. 2). Our regional phylogenetic analysis suggests that the introductions of the Alpha variant into the GCC originated from Europe [root-state posterior



**Figure 2.** Maximum clade credibility (MCC) phylogeny of the SARS-CoV-2 variants in the GCC countries (with related global isolates) between November 2020 and June 2021 and their demographic reconstruction of selected early introductions (encompassed by rectangles) in the Arabian Peninsula countries. (A, C, E, G, & H) Color of the branches indicate the most probable geographical location state of their descendent nodes and corresponds to the legend on the upper left of each tree. Circle sizes are proportional to the posterior node support ( $P > 0.5$  for larger circles and  $P < 0.5$  for smaller circles). The arrows point to the earliest probable introduction of the variant to the GCC countries from other regions with their inferred root state posterior probability (RSPP). The panels on the right (B, D, & F) represent Bayesian skygrid plots of the effective population size through time for the selected GCC clades within the phylogeny of the Alpha, Beta, and the Delta variants. The posterior median estimate is indicated by the dark colored line, and the light shaded areas correspond to the 95% high posterior density (HPD). The vertical dotted lines corresponds to the estimated time at which each variant transitioned from slow to fast population growth. (A-B) Alpha variant; (C-D) Beta variant; (E-F) represents Delta variant (G) represents Kappa variant; (H) represents Eta variant.

probability (RSPP) = 0.77], with strong posterior support, during July 2020 (95 per cent HPD [May 2020, August 2020]; Fig. 2A). Since the Alpha variant was introduced to the GCC, our demographic reconstruction of the dominant GCC clade showed plateauing effective population size after September 2020 followed by a decline in January 2021 (Fig. 2B). In contrast, our results showed that the introductions of the Beta and the Delta variants originated from Africa (RSPP = 0.76) in June 2020 (95 per cent HPD

[April 2020, September 2020]; Fig. 2D) and East Asia (RSPP = 0.98) in January 2021 (95 per cent HPD [December 2020, February 2021]; Fig. 2G). Our demographic reconstruction of the selected Beta clade dominant in the GCC post introduction showed two distinct peaks in January and May 2021 followed by notable declines in the effective populations (Fig. 2D). However, the effective population size of the Delta clade rapidly peaked and declined from June 2021 onward (Fig. 2F). All of the selected GCC clades partially mirrored our complete demographic construction (i.e. when sequences from all countries included, Supplementary Fig. S2A–C). Results of the mixed global and local demographic reconstructions for the effective population size through time for the five selected variants are illustrated in Supplementary Fig. S2. We found some support for the Kappa and the Eta variants originating from Oceania (RSPP = 0.71) in January 2021 (95 per cent HPD [November 2020, May 2021]) and Europe (RSPP = 0.70) in February (95 per cent HPD [August 2020, April 2021]), respectively, in which both of their estimated posterior statistical supports were substantially low ( $P < 0.5$ ; Fig. 2G, H). As the number of sequences from these variants was low (Tables S1 and S2), we were unable to reconstruct the demographic history of these variants within the GCC. However, we did detect substantially low effective population sizes for both the Kappa (Supplementary Fig. S2D) and the Eta (Supplementary Fig. S2E) variants with little sign of decline.

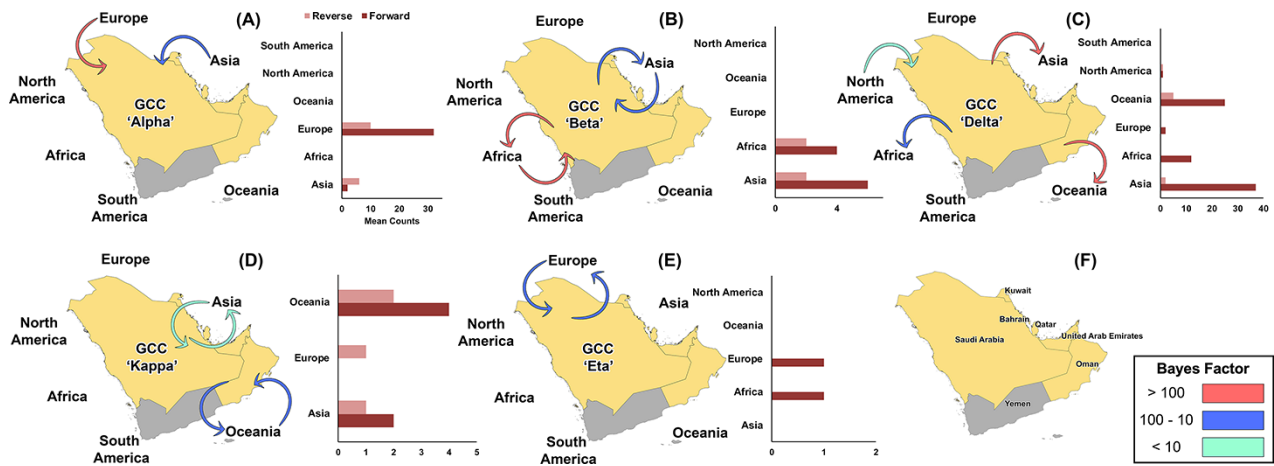
The BSSVS approach suggested that the most significant unidirectional dispersal route inferred for the Alpha variant was from Europe to the GCC ( $BF > 100$ ; Fig. 3A), whereas the least significant dispersal route was inferred from Asia to the GCC ( $BF < 100$ ; Fig. 3A). Additionally, Europe had the highest mean counts of relative forward transitions into the GCC (Forward = 32 vs Reverse = 10; Fig. 3A). Results of the Beta variant suggest that the most significant bidirectional viral dispersal routes were between the GCC and Africa ( $BFs > 100$ ; Fig. 3B), followed by Asia ( $BF < 100$ ; Fig. 3B). However, both Africa and Asia had notably higher forward than reverse transitions ( $MJs \geq 4$ ) in terms of viral jumps into the GCC (Fig. 3B). We found four significant unidirectional dispersal routes for the Delta variant (Fig. 3C), in which the most significant routes ( $BFs > 100$ ) were from the GCC to Asia and Oceania (Fig. 3C). However, the relative forward transitions from Asia and Oceania were substantially higher than those from the GCC ( $MJs > 25$ ; Fig. 3C). Further, the least significant dispersal routes inferred for the Delta variant were from the GCC to Africa and from North America to the GCC ( $BFs < 100$ ; Fig. 3C). Overall, we inferred least significant dispersal routes ( $BF < 100$ ) and low mean counts of viral jumps for the Kappa and the Eta variants (Fig. 3D, E). However, for both the Kappa and the Eta variants the mean count of forward transitions from different global regions into the GCC was substantially higher than from reverse transitions from the GCC (Fig. 3D, E).

## 4. Discussion

We used a rigorous comparative Bayesian phylodynamic statistical framework to untangle the evolutionary epidemiology of five SARS-CoV-2 variants in the GCC member states. We shed new insights into each variant's demographic history in parallel with the epidemic progression and geographic origins by quantifying their demographic history and migration routes between the GCC and other continents. Thus, our results represent a mixture of global and local demographic and phylogeographic reconstructions with special emphasis on the evolutionary characteristics of GCC viruses. These findings are not only critical for guiding current and future SARS-CoV-2 molecular surveillance programs

**Table 2.** Association indices and PSs for the discrete trait phylogeographic model of each selected variant.

Variant		AI <sup>a</sup>	95% CI	P-value	PS <sup>b</sup>	95% CI	P-value
Alpha	Observed	12.1	(10.4, 13.7)	<0.001	123.4	(117.0, 129.0)	<0.001
	Null	45.5	(43.7, 47.8)		282.4	(275.5, 289.9)	
Beta	Observed	3.4	(2.6, 4.2)	<0.001	38.0	(36.0, 40.0)	<0.001
	Null	19.7	(18.5, 21.0)		103.8	(101.4, 106.2)	
Delta	Observed	8.0	(6.7, 9.3)	0.009	90.9	(86.0, 96.0)	< 0.001
	Null	31.3	(29.8, 32.8)		178.9	(174.7, 182.6)	
Kappa	Observed	1.3	(1.0, 1.7)	0.004	9.7	(9.0, 11.0)	0.005
	Null	2.2	(1.7, 2.7)		15.0	(13.3, 16.5)	
Eta	Observed	0.5	(0.1, 0.9)	<0.001	6.1	(5.0, 7.0)	0.009
	Null	2.6	(2.1, 3.1)		14.3	(13.5, 14.9)	

<sup>a</sup>Association index (AI);<sup>b</sup>parsimony score (PS).

**Figure 3.** Regional exchange routes and posterior mean counts of the Markov jumps (MJ) over the phylogeny of the selected SARS-CoV-2 variants in the GCC countries between November 2020 and June 2021. Dispersal routes with none-zero rates were inferred using the Bayesian stochastic search variable selection (BSSVS) and were selected using Bayes factor comparisons. Significantly supported Dispersal routes (Bayes Factor > 3) are plotted. Arrows represent dispersal routes with the statistical significant posterior support (BSSVS BF > 3) as indicated by the legend on the lower left. The bar chart on the right of each map summarizes the total expected forward and reverse MJs between the GCC countries and other regions. (A) represents Alpha and was inferred from the asymmetric phylogeographic model with irreversible transitions; (B) represents Beta and was inferred from the symmetric phylogeographic model with reversible transitions; (C) represents Delta and was inferred from the asymmetric phylogeographic model with irreversible transitions; (D) represents Kappa and was inferred from the symmetric phylogeographic model with reversible transitions; (E) represents Eta and was inferred from the symmetric phylogeographic model with reversible transitions. (F) GCC countries' names. The maps were generated using by Q-GIS version 3.16 and SPREAD3 version 0.96.

globally and in the GCC but assist with the intervention efforts on this public health devastating pathogen.

Our phylogeographic analysis revealed that the Alpha, Beta, and Delta epidemic waves in the GCC member states resulted from multiple direct introductions from regions where the viruses were first identified (i.e. Europe, Africa, and East Asia, respectively; Fig. 2A, C, E) (Galloway et al. 2021; Kirola 2021; Mwenda et al. 2021; Yadav et al. 2021). Our findings support the notion that the evolution of SARS-CoV-2 emerging variants is strongly structured and maintained by their geographic origins (Hassan et al. 2021). Also, our results demonstrate that the increased number of reported cases in the regions of origin precipitates the introduction of each variant into the GCC region (Fig. 2A, C, E). For the Alpha variant, we found significant and intense (BF-BSSVS > 100, MJs > 30) unidirectional dispersal routes from Europe into the GCC region (Fig. 3A). In contrast, weakest and least supported dispersal routes were from Asia in the GCC (Fig. 3A). These findings are likely to reflect the effect of partial travel restrictions imposed by the GCC countries, in which citizens were allowed to return immediately to their countries while limiting the entries of expatriates and migrant workers (Hale et al. 2021). There is a

well-established preference for GCC citizens to travel to Europe for tourism, study, and medical care, and therefore, they may have been vectors of Alpha dispersal into the GCC when they were allowed to return. However, our analysis revealed significant bidirectional, although notably less intense, migration routes for the Beta variant between the GCC and Africa and Asia (Fig. 3B). This regional circulation might reflect the heterogeneous implementation of the travel restrictions across the GCC region. For example, Qatar (Benslimane et al. 2021) and UAE (Liu et al. 2021) both with major international airports allowed limited commercial flights mainly for foreign travelers throughout the second half of 2020 (Hale et al. 2021). Interestingly, our Delta variant results unveiled significant unidirectional dispersal from the GCC into Asia and Oceania, but the forward viral jumps from both regions were more intense than the GCC (Fig. 3C). The easing of non-pharmaceutical interventions in March and April 2021 may have facilitated the circulation and maintenance of the Delta variant in the region. The easing of movement restrictions for migrant workers may have been a driver of this pattern as they constitute a substantial portion of the total population in the GCC (International Labour Organization 2021). Migrant workers in the GCC largely originate

from East Asian countries such as India, Bangladesh, and Nepal, and the circulation between Asia and the GCC may be explained by the increased movement of these workers between countries after restrictions were lifted. The paucity of migration routes detected for the Kappa and Eta variants (Fig. 3D, E) highlights the reduced importance of these variants in the region. This finding is in line with what is known globally about the biology of these variants of interest. For example, the Kappa variant is known not to be as transmissible as Delta, and Eta is only found at low prevalence (Tao et al. 2021). Yet, the low prevalence of the Kappa and the Eta variants in the GCC (Supplementary Table S1 and S2) might significantly affect their posterior inferences.

Moreover, our phylogeographic results were used to inform our demographic analyses to help isolate the phylodynamic patterns at a regional scale. The robustness of our Bayesian Skygrid analysis was reflected by demonstrating how the effective population size of each targeted clade in the GCC was similar to the overall population size (Fig. 2B, D, F, Supplementary Fig. S2A–C). Further, we found a similar declining pattern after an increase in the population sizes of the GCC Alpha and the Beta clades that coincided with the temporal pattern of the reported cases (Supplementary Fig. S1). Yet, a second increase in the effective population size for the Beta variant in March and April 2021 (Fig. 2D) coincided with an increase in the number of sequenced isolates. This finding suggests the introduction of the second wave of Beta strains in the region. However, the effective population size of the Delta clade predominantly in the GCC remained high for approximately 4 months with a notable decline by the end of the study period (Fig. 2F) mirroring the temporal pattern of the overall reported cases in the region (Supplementary Fig. S1). These findings may indicate that the Delta variant was introduced into the GCC at later stages of the global Delta wave, since its emergence in East Asia in late 2020 [World Health Organization (WHO) 2021b]. This is not surprising, since most commercial flights between the GCC and East Asia were mostly restricted in early 2021 while the first sequenced isolates were reported in early March 2021 (GISAID.org) throughout the region in parallel to the beginning of easing such restrictions. Thus, given the high mutation rate of the Delta variant, the continued genomic surveillance should either detect a persistent decline in the genetic diversity of the circulating Delta viruses or introduction of further strains that could reaggregate the genetic diversity as inferred with the Beta variant (Fig. 2D). Such findings justify the allocation of additional resources for targeted genomic surveillance of the Delta variant over other selected variants during this specific period of time.

Our results also show the potential impact of pharmaceutical and non-pharmaceutical interventions on variant dynamics. Most GCC member states relaxed their non-pharmaceutical interventions at the beginning of 2021 following the large-scale rollout of vaccinations (Hale et al. 2021). At this time, our models show a rapid decline in the population sizes for Beta and Delta variants that may be attributed to this vaccination rollout (Fig. 2D, F). However, the slow decline of the genetic diversity of the Alpha variant in late January 2021 maybe attributed to the strict implementation of the non-pharmaceutical interventions and natural immunity resulting from the highly observed recovery rate in the region (Kojima and Klausner 2021; López et al. 2021). These findings are almost in agreement with the observed epidemic curve in the GCC region (Supplementary Fig. S1). It is worth noting that the waves dominated by the Beta and the Delta variant had coincided with an unusual, prolonged sandstorm that occurred during mid-March 2021 (Meo et al. 2021) and lasted for approximately 2 weeks

throughout the GCC region. This sandstorm resulted in a substantial increase in the air Particle Matter (PM) 2.5, subsequently leading to a sharp increase in the number of cases throughout the region (Supplementary Fig. S1). Also, it is worth that noting shorter duration (i.e. 2–4 days long) sand storms are frequent in the region between April and May of each year (Arishi 2021), and therefore, the air pollution accompanied by related anthropological behaviors could have aggravated and intensified the prolonged population growth of the local Delta variant (Fig. 2F), and subsequently, its transmissibility in the communities (Baron 2021). Furthermore, the epidemic curve of the GCC had two prominent peaks between May and July for both 2020 and 2021 (Supplementary Fig. S1), suggesting hints of seasonality in the transmission of the virus in the region (Fountain-Jones et al. 2020). Therefore, just like other respiratory diseases, air pollution and seasonality of the SARS-CoV-2 could be considered in future studies using generalized linear model-based phylodynamic methods (Tobias et al. 2021; Lemey et al. 2014).

Our study has limitations typically found in phylogenetic studies derived from publicly available genomic surveillance data. One important limitation is that our phylodynamic analyses used only sequences available at GISAID within the time frame of the present study. This restriction resulted in some biases in space and time for each variant. For example, Saudi Arabia and Oman were severely underrepresented in terms of the number of sequences and variants detected (see Fig. 1 and Supplementary Table S1). Furthermore, the availability of sequence metadata mainly related to travel history was extremely rare. Thus, such limitation prevented us from implementing the recently introduced phylodynamic method that accommodates individual travel history to generate less biased and more realistic posterior inferences (Lemey et al. 2020). Also, the notable inflation in the uncertainty estimates of substitution per site per year for each variant (Table 1) might be attributed to the short time frame of the sampling events (Supplementary Fig. S2) (Ghafari et al. 2022). That said, our phylodynamic inferences were based on all available genomic data collected by passive and active surveillance in the GCC region. Furthermore, our spatial discrete trait aggregation of the GCC countries into one unit might be helped in reducing the biases caused by missing information. For example, with the exception of the Beta variant, the demographic reconstruction results showed an early prolonged slow increase in the effective population size rather than a sharp increase even though most isolates were collected at a later stage of the study period (Supplementary Fig. S2A–E; Fig. 2B, F). Additionally, consistent declines in the population size were inferred, particularly for the Alpha and the Delta variants, which were also not affected by the later intensity of the sampling events (Fig. 2B, F). Further, for the Alpha, Beta, and Delta variants, we inferred the region from which they were first isolated, as the ancestral location of introduction, while the GCC sequences' number was substantially larger (Supplementary Table S2).

Saudi Arabia was the first country in the GCC that integrated phylodynamic methods in its genomic surveillance of SARS-CoV-2 (Mourier et al. 2022). However, their implementation focused on identifying mutations in the nucleocapsid (N) protein, followed by reconstructing population demographics and epidemic progression of earlier lineages (Mourier et al. 2022). Therefore, there were no rigorous attempts related to the direct implementation of phylogeographic reconstructions of multiple emerging variants in the region to the authors' knowledge. It is worth noting that revisiting the best fitting model for the selected variants

is critical when a new sequence dataset is made available, especially when sampling events are done over shorter time durations. This is because the evolutionary parameters of rapidly evolving pathogens like SARS-CoV-2 are time-dependent, in which best fitting node-age or/and branch rate might change over time with new sequence datasets, leading to different inferred posterior estimates (Ghafari et al. 2022). Furthermore, our analysis pipeline is focused on the comparative evolutionary characteristics of emerging variants to guide risk-based genomic surveillance efforts. Thus, in our GCC scenario, future genomic efforts should target the Delta variant and compare its evolutionary characteristics to the newly emerging variants, such as the Omicron variant. Also, if one decides to analyze multiple partial gene segments simultaneously, it is important to note that each gene segment might require different node-age or/and branch-rate model combinations, which might lead to similar or different posterior results.

## 5. Conclusions

Our study represents the first attempt in the GCC region and the Middle East to implement a rigorous Bayesian phylodynamic analytical pipeline to compare and target the evolutionary characteristics of emerging variants of SARS-CoV-2. We unveiled and compared the evolutionary demographics and phylogeographic history of the Alpha, Beta, Delta, Kappa, and Eta variants under pharmaceutical and non-pharmaceutical interventions in the GCC region. Our inferred phylodynamics for the Kappa and the Eta variants suggests that they no longer need to be targeted in genomic surveillance activities in the GCC. In contrast, the evolutionary characteristics of the Alpha, Beta, and Delta variants confirm their dominance in the GCC in the 2021 outbreaks and therefore need to be targeted, particularly the Delta variant. Due to the continued rapid evolution of SARS-CoV-2 and the emergence of new variants, increasing sequencing capacity and detailed reporting of their related metadata, particularly in undersampled geographical areas, is critical for improving the performance and accuracy of current genomic surveillance programs in the region. Therefore, our study highlights the urgent need to establish collaborative regional molecular surveillance programs to ensure effective decision making related to the allocation of intervention activities targeted toward the most relevant variants.

## Data availability

BEAST XML files and R-codes for the analysis of each variant are available at: <https://github.com/maalkhamis/Comparative-Phylodynamic-Models-for-SARS-CoV-2-in-the-GCC>.

## Supplementary data

Supplementary data is available at *Virus Evolution* online.

## Acknowledgements

This research was funded in part by the Kuwait Foundation for the Advancement of Science (Project No: 2020/1428). Accession numbers and their corresponding acknowledgment information for focal and context sequences in s1.focal.sequences.pdf and s2.context.sequences.pdf, respectively, in accordance with GISAID regulations.

**Conflict of interest:** None declared.

## References

- Al Awaidey, S. T. et al. (2021) 'COVID-19 in the Gulf Cooperation Council Member States: An Evidence of Effective Response', *Oman Medical Journal*, 36: e300.
- Al-Mahruqi, S. et al. (2021) 'Molecular Epidemiology of COVID-19 in Oman: A Molecular and Surveillance Study for the Early Transmission of COVID-19 in the Country', *International Journal of Infectious Diseases*, 104: 139–49.
- Andersen, K. G. et al. (2020) 'The Proximal Origin of SARS-CoV-2', *Nature Medicine*, 26: 450–2.
- Arishi, A. (2021) 'Classification of Sandstorms in Saudi Arabia', *Atmospheric and Climate Sciences*, 11: 177–93.
- Assiri, A. et al. (2013) 'Epidemiological, Demographic, and Clinical Characteristics of 47 Cases of Middle East Respiratory Syndrome Coronavirus Disease from Saudi Arabia: A Descriptive Study', *The Lancet Infectious Diseases*, 13: 752–61.
- Ayres, D. L. et al. (2019) 'BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics', *Systematic Biology*, 68: 1052–61.
- Baele, G. et al. (2012) 'Improving the Accuracy of Demographic and Molecular Clock Model Comparison while Accommodating Phylogenetic Uncertainty', *Molecular Biology and Evolution*, 29: 2157–67.
- Baron, Y. M. (2021) 'Are There Medium to Short-term Multifaceted Effects of the Airborne Pollutant PM(2.5) Determining the Emergence of SARS-CoV-2 Variants?' *Medical Hypotheses*, 158: 110718.
- Benslimane, F. M. et al. (2021) 'One Year of SARS-CoV-2: Genomic Characterization of COVID-19 Outbreak in Qatar', *Frontiers in Cellular and Infection Microbiology*, 11: 768883.
- Benvenuto, D. et al. (2020) 'The Global Spread of 2019-nCoV: A Molecular Evolutionary Analysis', *Pathogens and Global Health*, 114: 64–7.
- Bielejec, F. et al. (2016) 'Spread3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes', *Molecular Biology and Evolution*, 33: 2167–9.
- Bindayna, K. M., Deifalla, A., and Mokbel, H. E. M. (2021) 'Identification of E484K and Other Novel SARS-COV-2 Variants from the Kingdom of Bahrain', *Microbial Pathogenesis*, 157: 104955.
- Bolyen, E. et al. (2020) 'Reproducibly Sampling SARS-CoV-2 Genomes across Time, Geography, and Viral Diversity', *F1000Research*, 9: 657.
- Butera, Y. et al. (2021) 'Genomic Sequencing of SARS-CoV-2 in Rwanda Reveals the Importance of Incoming Travelers on Lineage Diversity', *Nature Communications*, 12: 5705.
- De Maio, N. et al. (2015) 'New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation', *PLoS Genetics*, 11: e1005421.
- Farah, S. et al. (2020) 'Phylogenomics and Phylodynamics of SARS-CoV-2 Genomes Retrieved from India', *Future Virology*, 15: 747–53.
- Farr, W. M., Mandel, I., and Stevens, D. (2015) 'An Efficient Interpolation Technique for Jump Proposals in Reversible-Jump Markov Chain Monte Carlo Calculations', *Royal Society Open Science*, 2: 601.
- Fountain-Jones, N. M. et al. (2020) 'Emerging Phylogenetic Structure of the SARS-CoV-2 Pandemic', *Virus Evolution*, 6: veaa082.
- Galloway, S. E. et al. (2021) 'Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021', *MMWR Morbidity and Mortality Weekly Report*, 70: 95–9.
- Geoghegan, J. L. et al. (2020) 'Genomic Epidemiology Reveals Transmission Patterns and Dynamics of SARS-CoV-2 in Aotearoa New Zealand', *Nature Communications*, 11: 6351.
- Ghafari, M. et al. (2022) 'Purifying Selection Determines the Short-Term Time Dependency of Evolutionary Rates in SARS-CoV-2 and pH1N1 Influenza1', *Molecular Biology and Evolution*, 39: msac009.



- Gill, M. S. et al. (2013) 'Improving Bayesian Population Dynamics Inference: A Coalescent-based Model for Multiple Loci', *Molecular Biology and Evolution*, 30: 713–24.
- Giovanetti, M. et al. (2020) 'The First Two Cases of 2019-nCoV in Italy: Where They Come From?', *Journal of Medical Virology*, 92: 518–21.
- Gomez, C. E., Perdiguero, B., and Esteban, M. (2021) 'Emerging SARS-CoV-2 Variants and Impact in Global Vaccination Programs against SARS-CoV-2/COVID-19', *Vaccines (Basel)*, 9: 1405–14.
- Griffiths, R. C., and Tavaré, S. (1994) 'Ancestral Inference in Population Genetics', *Statistical Science*, 9: 307–19.
- Grubaugh, N. D. et al. (2019) 'Tracking Virus Outbreaks in the Twenty-first Century', *Nature Microbiology*, 4: 10–9.
- Hadfield, J. et al. (2018) 'NextStrain: Real-time Tracking of Pathogen Evolution', *Bioinformatics*, 34: 4121–3.
- Hale, T. et al. (2021) 'A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)', *Nature Human Behaviour*, 5: 529–38.
- Hassan, S. S. et al. (2021) 'Implications Derived from S-Protein Variants of SARS-CoV-2 from Six Continents', *International Journal of Biological Macromolecules*, 191: 934–55.
- International Labour Organization. (2021), *Labour Migration (Arab States)*. International Labour Organization. <<https://www.ilo.org/beirut/areasofwork/labour-migration/lang-en/index.htm>> accessed 15 Nov 2021.
- Jeewandara, C. et al. (2021) 'Genomic and Epidemiological Analysis of SARS-CoV-2 Viruses in Sri Lanka', *Frontiers in Microbiology*, 12: 722838.
- Kalyaanamoorthy, S. et al. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*, 14: 587–9.
- Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.
- Kingman, J. F. C. (1982) 'On the Genealogy of Large Populations', *Journal of Applied Probability*, 19: 27–43.
- Kirola, L. (2021) 'Genetic Emergence of B.1.617.2 In COVID-19', *New Microbes and New Infections*, 43: 100929.
- Kojima, N., and Klausner, J. D. (2021) 'Protective Immunity after Recovery from SARS-CoV-2 Infection', *The Lancet Infectious Diseases*. [10.1016/S1473-3099\(21\)00676-9](https://doi.org/10.1016/S1473-3099(21)00676-9).
- Larsson, A. (2014) 'AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets', *Bioinformatics*, 30: 3276–8.
- Lemey, P. et al. (2020) 'Accommodating Individual Travel History and Unsourced Diversity in Bayesian Phylogeographic Inference of SARS-CoV-2', *Nature Communications*, 11: 564–70.
- et al. (2014) 'Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2', *PLoS Pathogens*, 10: e1003932.
- et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.
- Liu, R. et al. (2021) 'Genomic Epidemiology of SARS-CoV-2 in the UAE Reveals Novel Virus Mutation, Patterns of Co-Infection and Tissue Specific Host Immune Response', *Scientific Reports*, 11: 2196–203.
- López, M. G. et al. (2021) 'The First Wave of the COVID-19 Epidemic in Spain Was Associated with Early Introductions and Fast Spread of a Dominating Genetic Variant', *Nature Genetics*, 53: 1405–14.
- Lycett, S. J. et al. (2021) 'Epidemic Waves of COVID-19 in Scotland: A Genomic Perspective on the Impact of the Introduction and Relaxation of Lockdown on SARS-CoV-2', *medRxiv*. [10.1101/2021.01.08.20248677](https://doi.org/10.1101/2021.01.08.20248677).
- Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.
- Meo, S. A. et al. (2021) 'Sandstorm and Its Effect on Particulate Matter PM 2.5, Carbon Monoxide, Nitrogen Dioxide, Ozone Pollutants and SARS-CoV-2 Cases and Deaths', *Science of the Total Environment*, 795: 148764.
- Miller, D. et al. (2020) 'Full Genome Viral Sequences Inform Patterns of SARS-CoV-2 Spread into and within Israel', *Nature Communications*, 11: 5518.
- Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.
- Morel, B. et al. (2021) 'Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult', *Molecular Biology and Evolution*, 38: 1777–91.
- Mourier, T. et al. (2022) 'SARS-CoV-2 Genomes from Saudi Arabia Implicate Nucleocapsid Mutations in Host Response and Increased Viral Load', *Nature Communications*, 13: 601.
- Mwenda, M. et al. (2021) 'Detection of B.1.351 SARS-CoV-2 Variant Strain - Zambia, December 2020', *MMWR Morbidity and Mortality Weekly Report*, 70: 280–2.
- Nabil, B., Sabrina, B., and Abdelhakim, B. (2020) 'Transmission Route and Introduction of Pandemic SARS-CoV-2 between China, Italy, and Spain', *Journal of Medical Virology*, 93: 564–8.
- Nandkeolyar, K. (2020), *Coronavirus in UAE: Four of a Family Infected* <<https://gulfnnews.com/uae/health/coronavirus-in-uae-four-of-a-family-infected-1.1580273983681>> accessed 20 Oct 2021.
- Parker, J., Rambaut, A., and Pybus, O. G. (2008) 'Correlating Viral Phenotypes with Phylogeny: Accounting for Phylogenetic Uncertainty', *Infection, Genetics and Evolution*, 8: 239–46.
- Pybus, O. G., and Rambaut, A. (2009) 'Evolutionary Analysis of the Dynamics of Viral Infectious Disease', *Nature Reviews Genetics*, 10: 540–50.
- Rambaut, A. (2018), *FigTree*. In (Version 1.4) <<http://tree.bio.ed.ac.uk/software/figtree/>> accessed 12 Dec 2021.
- et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.
- et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: 1.
- Sagulenko, P., Puller, V., and Neher, R. A. (2018) 'TreeTime: Maximum-likelihood Phylodynamic Analysis', *Virus Evolution*, 4: vex042.
- Seemann, T. et al. (2020) 'Tracking the COVID-19 Pandemic in Australia Using Genomics', *Nature Communications*, 11: 4376.
- Suchard, M. A. et al. (2018) 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10', *Virus Evolution*, 4: 815–21.
- Tao, K. et al. (2021) 'The Biological and Clinical Significance of Emerging SARS-CoV-2 Variants', *Nature Reviews Genetics*, 22: 757–73.
- Tayoun, A. A. et al. (2020) 'Multiple Early Introductions of SARS-CoV-2 into a Global Travel Hub in the Middle East', *Scientific Reports*, 10: 17720.
- Tobias, A. et al. (2021) 'Can SARS-CoV-2 Global Seasonality Be Determined after One Year of Pandemic?', *Environmental Epidemiology (Philadelphia, Pa.)*, 5: e146–e146.
- WHO Emergency Committee. Statement on the Second Meeting of the International Health Regulations. (2005), *Emergency Committee regarding the Outbreak of Novel Coronavirus (COVID-19)*. Geneva: WHO, 2020. <[https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))> accessed 1 Feb 2020.
- World Health Organization (WHO). (2021a), *Coronavirus Disease (COVID-19) Dashboard*. <<https://covid19.who.int>> accessed 10 Oct 2021.
- (2021b), *Tacking SARS-CoV-2 Variants*. <<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>> accessed 20 Dec 2021.

- Worobey, M. et al. (2020) 'The Emergence of SARS-CoV-2 in Europe and North America', *Science*, 370: 564–70.
- Yadav, P. D. et al. (2021) 'Neutralization of Variant under Investigation B.1.617 with Sera of BBV152 Vaccinees', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 74: 366–8.
- Zhou, H. et al. (2020) 'A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein', *Current Biology*, 30: 2196–203 e2193.
- Zhu, N. et al. (2020) 'A Novel Coronavirus from Patients with Pneumonia in China, 2019', *New England Journal of Medicine*, 382: 727–33.