Cryptic spread of the SARS-CoV-2 Delta variant undermined targeted travel restrictions in Iraq

Praneeth Gangavarapu[1], Edyth Parker[1], Mark Zeller[1], Jassim Mohammed Abdo[2], Refugio Robles-Sikisaka[1], Ezra Kurzban[1], Sarah Perkins[1], Madison Schwab[1], Kelly Nguyen[1], Michelle McGraw[1], Karthik Ramesh[1], Catelyn Anderson[1], Narin A. Rasheed[2], Renas Husain Isa[2], Sherzad Majeed Taher[2], Kristian Andersen[1]

[1] *Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA*
[2] *University of Duhok COVID-19 Center for Research and Diagnosis; University of Duhok; Kurdistan Region; Iraq*

Abstract

Genomic epidemiology is the use of pathogen genomic data to study the spread of a disease in a population. In this study, we investigated the spread of the SARS-CoV-2 virus in Iraq and the effectiveness of targeted travel restrictions implemented by the Iraqi government. Using Bayesian phylogenetic methods, we reconstructed a time-resolved phylogeny and the geographic spread of the virus between Iraq and other regions in the world. We found that the majority of the introductions into Iraq originated from Southern Asia and the Middle East. We also found that the first introductions of the Delta variant into Iraq occurred months before travel restrictions were implemented and that these introductions resulted in ongoing community transmission. We show that new virus introductions were not effectively halted during the period with travel restrictions, with the majority of introductions occurring from Western and Southern Asia. Our findings suggest that travel restrictions targeting a single source country are unlikely to work for a rapidly spreading pathogen in a highly connected world with limited genomic surveillance.

Introduction

The global expansion of genomic surveillance infrastructure during the COVID-19 pandemic has enabled the characterization of pathogen transmission dynamics at unprecedented temporal, geographic and individual scales.[1–3] This has empowered studies investigating the local and global drivers of SARS-CoV-2 transmission dynamics as well as those assessing the efficiency of mitigation efforts, such as travel or movement restrictions.[2,4] Regional connectivity has been emphasized as a major driver of SARS-CoV-2 transmission in numerous studies across Europe and Africa.[2,3,5–7] However, only a single study from Jordan has investigated transmission drivers in the Middle East, which showed that regional connectivity, specifically shorter distance land-based travel, disproportionately drove bidirectional

transmission of SARS-CoV-2 in the Middle East when travel restrictions were in effect.[8] A more extensive investigation of SARS-CoV-2 genomic epidemiology in the Middle East is limited by severe undersampling in the region, since every country in the Middle East sequenced less than one percent of all SARS-CoV-2 cases across 2021.[9] Without additional sampling, the generalizability of drivers of transmission identified in other regions remains unclear.[8]

The Delta variant was first detected in India in March 2021, though it is estimated to have emerged in mid-October 2020.[4] The variant circulated cryptically for 6 months, spreading globally to 97 countries by May 2021, and seeding over 121 sublineages.[10] Iraq, like many other countries, implemented travel restrictions to and from India during its Delta-driven epidemic wave to mitigate the introduction of the variant.[11] Despite an absolute travel restriction executed on 27 April 2021 after the first case detection, Iraq experienced a Delta-driven outbreak wave from June to September 2021. Though the targeted travel restriction was unsuccessful, the predominant sources, other than India, and timing of introductions in Iraq during the Delta wave remains unclear. In particular, due to limited sampling, it is unclear whether community transmission was already established in Iraq before travel restrictions were implemented, or whether the virus was introduced during the period with travel restrictions via other countries. It is also unclear what role regional connectivity played in the source-sink dynamics during the Delta wave, as Iraq is highly socio-politically and economically connected to its neighbors.

In this study, we characterized the transmission dynamics of SARS-CoV-2 in Iraq across the Delta wave from June to December 2021. We focused our analyses on the Delta wave to test the effectiveness of a travel restriction targeting India alone, spanning from 27 April 2021 to 6 September 2021. We reconstructed the temporal and geographic spread of the virus using Bayesian phylogeography and show that the Delta variant was first introduced in early February 2021, three months before travel restrictions were implemented. Notably, we found that six of the seven large circulating transmission chains were introduced before the travel restrictions were in place. This suggests that community transmission was already ongoing when travel restrictions went into effect; thus, restrictions were implemented too late to mitigate or delay the spread of the Delta variant in Iraq. We further show that many early introductions originated from neighboring countries in the Middle East, indicating that the travel restrictions targeting India were unlikely to be successful for a rapidly spreading variant in a highly connected world.

Results

The majority of the Delta variant introductions occurred after travel restrictions were implemented

To characterize the number of viral introductions and exports for Iraq, we generated a total of 535 sequences from June 2021 to March 2021 from samples collected from hospitalized patients in Duhok, Iraq (**Fig. 1A**). Our study period covered the decline of the Alpha variant wave (n=12 sequences), the emergence and decline of the Delta variant wave (n=343 sequences), and the beginning of the Omicron variant wave (n=180 sequences) (**Fig. 1B**). In this study, we focus on the subset of Delta sequences to investigate the effect of restricted travel to and from India. This travel restriction was enacted by the Iraqi government from 27 April 2021 to 6 September 2021 in response to the Delta wave.

We first detected the Delta variant in Iraq on 8 June 2021. The variant subsequently drove the nation's third epidemic wave from July to December 2021 (**Fig. 1B, 1C**). We found that the ancestral B.1.617.2 lineage and sublineage AY.33 dominated the wave, with several other sublineages co-circulating after the ancestral lineages' long cryptic circulation and global expansion (**Fig. 1C**).[4]

It is currently unknown how the travel restriction on India may have delayed or reduced the number of introductions over time, and how the source countries of viral introductions changed over time in response to the travel restriction. To elucidate the impact of the travel restriction, we used Bayesian phylogeography to reconstruct the spatiotemporal spread of the virus before, during, and after the targeted travel restriction. We found that Southern Asia and the Middle East were the major sources of introductions into Iraq, with very few introductions coming from other regions (**Fig. 1D**). Overall, we estimated a total of 12 (95% HPD: 7 - 16) and 7 (95% HPD: 3 - 14) introductions from Southern Asia and the Middle East, respectively (**Fig. 1D**). This is consistent with studies demonstrating that the Delta variant emerged in India in mid-October and expanded globally by early May 2021.[4] The proportionally high number of introductions from the Middle East strongly suggests that travel restrictions targeting India likely shifted the profile of introductions to other regions.

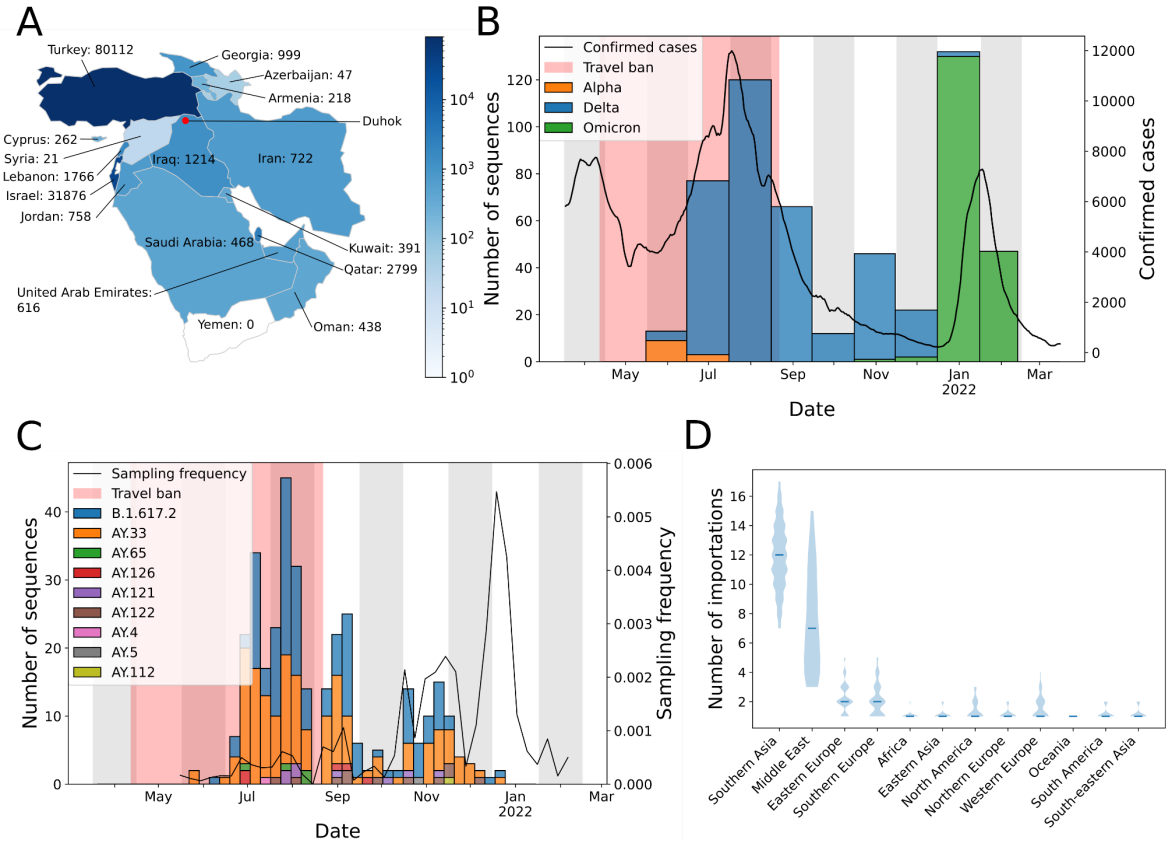

Figure 1. SARS-CoV-2 epidemiology in Iraq. (**A**) Geographic distribution of sequences collected in 2021 in Iraq and its neighboring countries (colored using log scale). Duhok, Iraq, is highlighted with a red dot. (**B**) Number of sequences collected by month (bar graph, left axis) and the number of COVID-19 cases (7-day rolling average) (line graph, right axis). (**C**) Number of Delta variant sequences collected binned by week (bar graph, left axis) and the sampling fraction of collected sequences for Iraq (7-day rolling average) (line graph, right axis). (**D**) Total estimated number of importations into Iraq from various source regions (95% HPD with median).

We estimated that the first sampled introduction of the Delta variant occurred in early February 2021, originating in Southern Asia (95% HPD: 28 December 2020 to 13 March 2021) (**Fig. 2A**). This is approximately one month prior to the first detection of Delta in India and three months after its estimated emergence in India in October 2020. This further strengthens current evidence that the Delta variant was spreading cryptically around the world before its detection in March 2021 in India.[4] The first case of the Delta variant in Iraq was confirmed on 27 April 2021.

Although we cannot rule out that the Delta variant was cryptically circulating in Iraq for almost three months prior to detection, the number of COVID-19 cases only increased sharply from June onwards suggesting a delay between the first introductions and widespread community transmission (**Fig. 1B**). However, it is impossible to precisely pinpoint when community transmission was established without more comprehensive sampling. Our first introduction from outside of Southern Asia originated from the Middle East in late March (95% HPD: 27 January 2021 - 9 July 2021) (**Fig. 2A**). Thus, we estimated that the first introductions from both Southern Asia and the Middle East occurred at least two months before any travel restrictions were implemented. This indicates that the single-source travel

restriction was implemented too late relative to Delta's cryptic emergence in India to be effective. It also suggests that a single-source travel restriction is unlikely to substantially delay the introduction of a rapidly spreading variant, as introductions from countries outside of Southern Asia occurred very shortly after those from India.

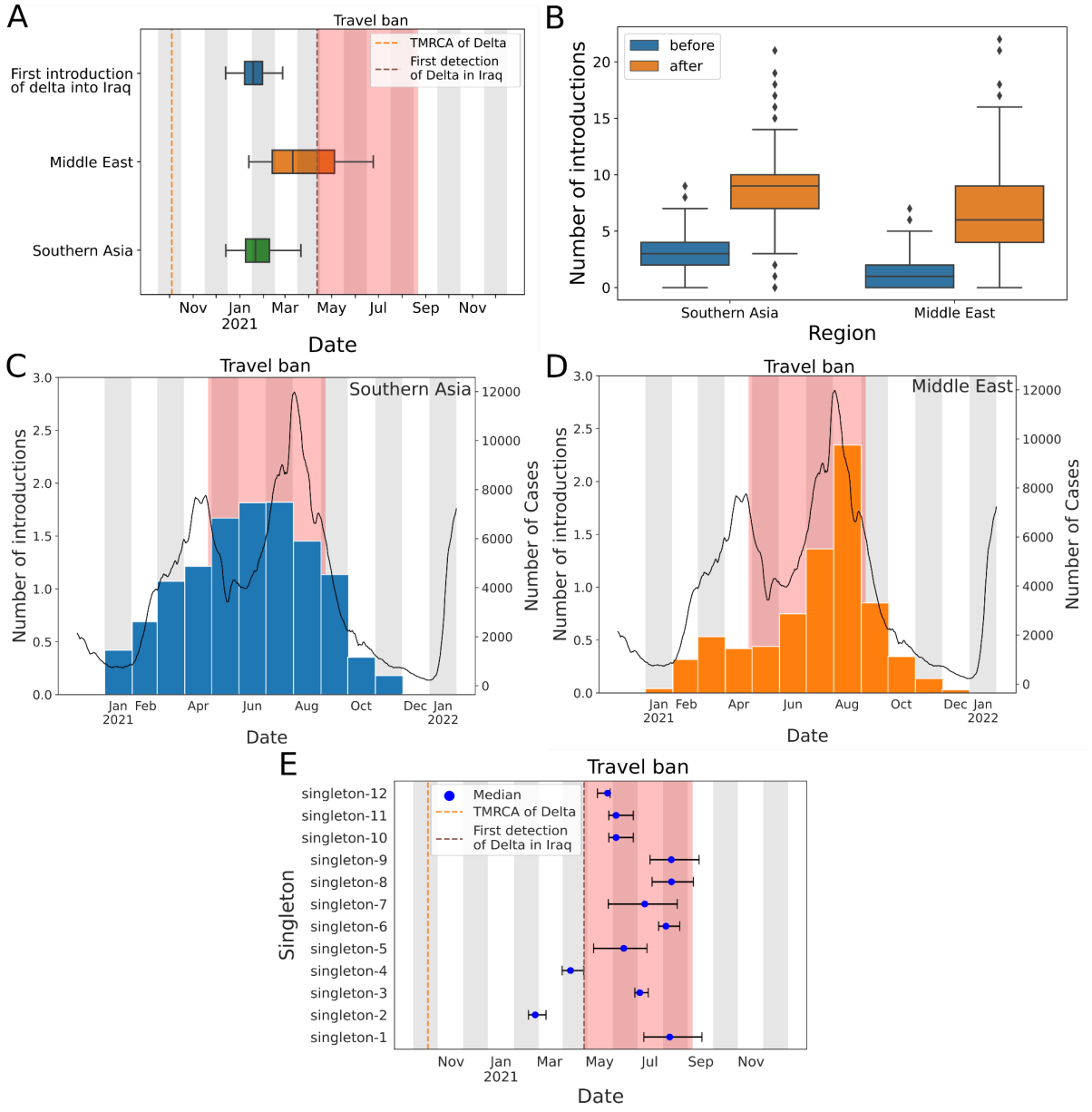

Figure 2. Effect of travel restrictions on viral emergence in Iraq. (**A**) Estimated first introduction into Iraq, first introduction from the Middle East and Southern Asia into Iraq. (**B**) Shift in the number of introductions before and after single-source travel restriction (27 April 2021). (**C**) Mean number of jumps from Southern Asia over time (bar graph, left axis) and the number of cases (7-day rolling average) (line graph, right axis). (**D**) Mean number of jumps from the Middle East over time (bar graph, left axis) and the number of cases (7-day rolling average) (line graph, right axis). (**E**) Estimated date of introduction of single introductions into Iraq.

We estimated that the majority of sampled introductions from Southern Asia and the Middle East, across the entire Delta wave, occurred after travel restrictions were in place (27 April 2021 to 6 September 2021) (**Fig. 2B**). The number of introductions from Southern Asia and the Middle East peaked in June-July and August when the

epidemic was already growing exponentially (**Fig. 2C-D, Fig. 1B**). However, this peak was largely driven by a high number of single introductions, and we estimated that only a single transmission chain emerged after travel restrictions were in place (**Fig. 2E**). Together with the exponential epidemic growth observed at the time **(Fig. 1B)**, this suggests that Delta's growth at the time was driven by community transmission rather than high rates of introductions.

Community transmission of the Delta variant was established months before the travel restrictions were implemented

We estimated that the first introduction occurred three months before travel restrictions were instated. However, it is unclear whether these introductions were contained or resulted in wider community transmission at the time. COVID-19 cases in the Delta-driven wave increased exponentially from June 2021 onwards, suggesting community transmission was delayed by a few months (**Fig. 1A**). To investigate whether the onset of community transmission was delayed by the travel restriction on India, we analyzed the timing and dynamics of the transmission chains in our dataset (**Fig. 3A**). We defined a transmission chain as an introduction that was responsible for 2 or more sampled cases whose descendants are from the same region.

We sampled seven transmission chains of ranging sizes (TC-1-7) (**Fig. 3A**), and found that five of the seven transmission chains were originally introduced from Southern Asia (TC-1, TC-2, TC-4, TC-6, and TC-7), with the remaining two introduced from the Middle East (TC-3 and TC-5) (**Fig. 3A**). All of the transmission chains persisted across the entire Delta wave. The transmission chain with the longest persistence was TC-7, sampled across 309 days, whereas the largest transmission chain was TC-1 with 118 sampled cases (35% of the sampled sequences) (**Fig. 3A**).

Interestingly, we found that six out of the seven sampled transmission chains were introduced and established before the travel restriction was implemented in late April (**Fig. 3B**). We found that AY.33 sublineage transmission chains TC-5, TC-6, TC-7 were introduced first, with their first sampled transmissions in Iraq occurring as early as February. We found that ancestral lineage B.1.617.2 (TC-1 and TC-2) and AY.121 (TC-4) transmission chains were also imported and established months before travel restrictions were in place **(Fig. 3B)**.

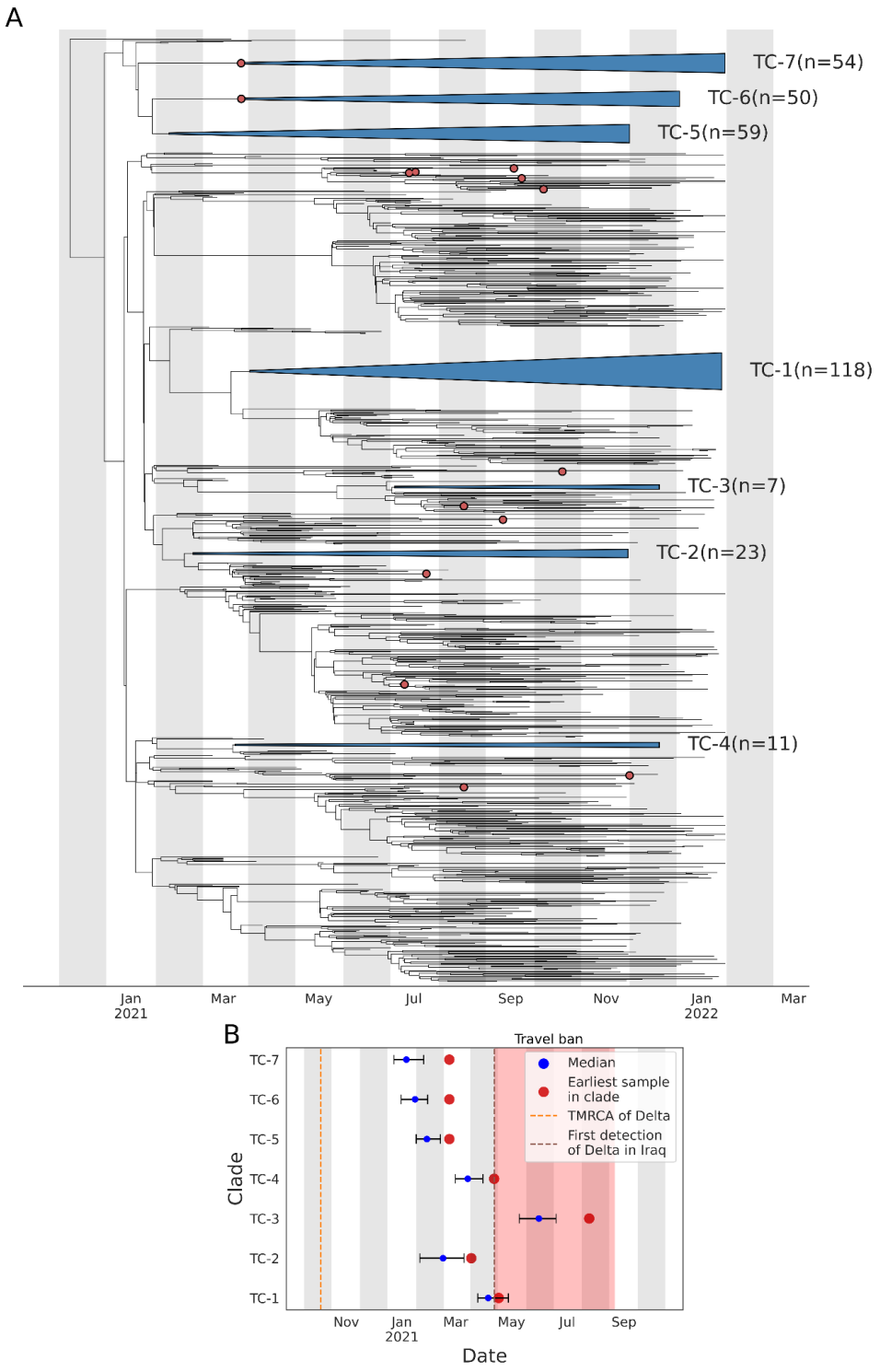

Figure 3. Introduction and dynamics of transmission chains. (A) A representative time-calibrated phylogeny. Transmission chains (TC-1, TC-2, TC-3, TC-4, TC-5, TC-6, TC-7) are collapsed into blue triangles. Red dots represent sampled singleton sequences from Iraq. (**B**) Estimated date of introduction of transmission chains.

To better understand the onset of community transmission, we analyzed the discrepancy between the date of emergence of community transmission chains and the date of detection. We observed the greatest delay in detection, 55 days, for the TC-3 transmission chain, followed by TC-7 and TC-6 at 47 and 37 days, respectively.

TC-1, the largest chain, had the shortest detection lag at 11 days (**Fig. 4A-B**). Overall, we observed a lower bound surveillance lag of one month between introduction and first detection for most transmission chains.

Discussion

In this study, we used phylogeographic reconstructions to investigate the effectiveness of a source-country targeted travel restriction instated in Iraq after the emergence of the Delta variant. We specifically questioned whether the restriction reduced or delayed the introduction of the variant into Iraq and how it shifted the source-sink dynamics of viral imports from the source country to regional countries. We found that the Delta variant was introduced three months before travel restrictions were implemented, with community transmission estimated to have been established approximately two months prior to the restrictions as well. This finding implies that travel restrictions were implemented far too late relative to Delta's cryptic, undersampled emergence in India to substantially reduce or delay spread to Iraq. Though the majority of our introductions originated in Southern Asia, we also found a number of early introductions from the Middle East. Thus, in the case of a rapidly dispersing pathogen, single-country travel restrictions are likely ineffective when surveillance is limited due to potential introductions via third countries.

We observed a lower bound one-month delay between the first introduction and detection as well as the emergence and detection of transmission chains after the onset of community transmission. This highlights the need for comprehensive, real-time surveillance to better inform public health guidelines and country-level policies, as travel restrictions may not have been imposed if ongoing transmission was detected in a timely manner.

Our findings should be contextualized by our limited sample size, which represents less than 0.01% of cases within the study period. Therefore, the number of single introductions and the size of the transmission chains that we report may be severely underestimated.[2] Moreover, regional introductions from neighboring and regional Middle Eastern countries are also likely to be underestimated due to undersampling of those regions.[8] Importantly, our sampling was concentrated in Duhok, a city in northern Iraq close to the border with Turkey and Syria, and may not be representative of population-level transmission dynamics in the rest of the country (e.g., more introductions from neighboring countries may have been observed due to proximity to the border). However, it is highly likely that earlier introductions and community transmission occurred unsampled outside of Duhok.

Methods

Sample collection

SARS-CoV-2 samples were collected from routine diagnostic qPCR tests performed by the University of Duhok COVID-19 center, Duhok, Iraq. To ensure a high success rate, we only sequenced positive qPCR samples with a Ct value less than 30. In total, we collected and sequenced 535 SARS-CoV-2 samples, effectively doubling the number of sequences from Iraq. The samples were collected from June 2021 until February 2022, spanning the beginning of the B.1.1.7 wave, the entire Delta wave (June to December 2021), and the beginning of the Omicron wave. All sequence data generated for this study have been made publicly available on GISAID (GISAID accession IDs are available on https://github.com/andersen-lab/HCoV-19-Genomics).[12]

SARS-CoV-2 whole-genome sequencing

Samples were shipped to the Andersen lab at Scripps Research on dry ice, and sequenced using an amplicon-based sequencing assay using overlapping ~250nt amplicons.[13] Briefly, viral RNA was extracted using the Mag-Bind Viral DNA/RNA kit (Omega Bio-tek) according to the manufacturer's instructions. SARS-CoV-2 RNA (2 µL) was reverse transcribed with SuperScript IV VILO (ThermoFisher Scientific). The virus cDNA was amplified in two multiplexed PCR reactions using Q5 DNA High-fidelity Polymerase (New England Biolabs). Following an AMPureXP bead (Beckman Coulter) purification of the combined PCR products, the samples were barcoded in another PCR reaction using barcoded Illumina adapters as primers. The libraries were again purified with AMPureXP beads and quantified using the Qubit High Sensitivity DNA assay kit (Invitrogen) and Tapestation D5000 tape (Agilent). The individual libraries were then normalized and pooled in equimolar amounts at 1.5 nM. The 1.5 nM library pool was sequenced using a NovaSeq 6000 SP Reagent Kit v1.5 (300 cycles). Consensus sequences were assembled using an in-house Snakemake pipeline with bwa-mem and iVar v1.2.2.[14–16]

Downsampling strategies and dataset curation

To generate a representative but downsampled dataset for phylogenetic analysis, we first downloaded all publicly available sequences of the lineages B.1.617.2, AY.4, AY.5, AY.33, AY.112, AY.121, AY.122, and AY.126 up until May 2022 from GISAID. Sequences with low coverage (>1% N's and >0.5% unique amino acid mutations) or with incomplete collection dates were excluded. Sequences were randomly subsampled proportional to the weekly number of new positive COVID-19 cases in each country. The number of new COVID cases per country was obtained from the COVID-19 dataset generated by Our World in Data.[17]

Next, sequences that deviated more than three interquartile ranges from the molecular clock were excluded. For this, the downsampled sequence dataset was

aligned to reference genome Hu-1 (GenBank accession number: MN908947.3) using MAFFTv7.505, and the 5' and 3' untranslated regions were trimmed, and all problematic sites in the alignment were masked.[18,19] The masked, aligned dataset was then used to reconstruct a maximum likelihood phylogenetic tree with IQTREE-2 using a ModelFinder and used TreeTime to generate a root-to-tip plot.[20–22]

Finally, sequences clustering within a genetically divergent subclade of the B.1.617.2 lineage were removed. The final downsampled and curated dataset contained 1152 sequences. Pangolin was used to assign a lineage to each sequence in the alignment.[23] A list of GISAID ids for our final downsampled sequence dataset can be found using the following GISAID EPI_SET Id: EPI_SET_221201da.[24]

Phylogenetic analyses

BEAST 1.10.5 was used to reconstruct a time-scaled phylogeny using a HKY substitution model with a gamma-distributed rate variation among sites, a strict clock with a lognormal prior, an exponential growth coalescent tree prior, and a flexible Skygrid coalescent prior with grid points every two weeks.[25] The Markov Chain Monte Carlo chain was run for 200 million states with default BEAGLE computational library parameters, and trees were sampled every 10,000 steps.[25] Tracer v1.7.2 was used to assess the convergence (Effective Sample Size >100) of the chain after removing a 10% burn-in.[26] We used TreeAnnotator 1.10 to obtain the maximum clade credibility phylogenetic tree and used the Baltic library to visualize the tree.[27,28]

To reconstruct the phylogeographic history, we performed an asymmetric discrete state analysis to reconstruct transmission dynamics between Iraq and geographic regions (as defined by the UN) by summarizing the start and end location of Markov jumps across the full posterior of all trees using TreeMarkovJumpHistoryAnalyzer.[29,30]

References

1. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).
2. du Plessis, L. *et al.* Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
3. Tegally, H. *et al.* Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.* **27**, 440–446 (2021).
4. McCrone, J. T. *et al.* Context-specific emergence and growth of the SARS-CoV-2 Delta variant. *Nature* **610**, 154–160 (2022).
5. Dudas, G. *et al.* Emergence and spread of SARS-CoV-2 lineage B.1.620 with variant of concern-like mutations and deletions. *Nat. Commun.* **12**, 5769 (2021).
6. Butera, Y. *et al.* Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance of incoming travelers on lineage diversity. *Nature Communications* vol. 12 Preprint at https://doi.org/10.1038/s41467-021-25985-7 (2021).
7. Nadeau, S. A. *et al.* Swiss public health measures associated with reduced SARS-CoV-2 transmission using genome data. *Sci. Transl. Med.* eabn7979 (2022).

8. Parker, E. *et al.* Regional connectivity drove bidirectional transmission of SARS-CoV-2 in the Middle East during travel restrictions. *Nat. Commun.* **13**, 4784 (2022).
9. Hughes, L. *et al.* Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Res Sq* (2022) doi:10.21203/rs.3.rs-1723829/v1.
10. O'Toole, Á. *et al.* Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. *Wellcome Open Res* **6**, 121 (2021).
11. Hale, T. *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* **5**, 529–538 (2021).
12. Khare, S. *et al.* GISAID's Role in Pandemic Response. *China CDC Wkly* **3**, 1049–1051 (2021).
13. Home. *SEARCH* https://searchcovid.info/ (2020).
14. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).
15. Köster, J. & Rahmann, S. Building and documenting workflows with python-based snakemake. in *German Conference on Bioinformatics 2012* 56 (Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany, 2012).
16. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).
17. *covid-19-data: Data on COVID-19 (coronavirus) cases, deaths, hospitalizations, tests • All countries • Updated daily by Our World in Data*. (Github).
18. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
19. *ProblematicSites_SARS-CoV2*. (Github).
20. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
21. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
22. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol 4: vex042. Preprint at (2018).
23. O'Toole, Á. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evolution, 7 (2), veab064. Preprint at (2021).
24. [No title]. https://epicov.org/epi3/epi_set/221201da.
25. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016 (2018).
26. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
27. Dudas, G. *baltic: baltic - backronymed adaptable lightweight tree import code for molecular phylogeny manipulation, analysis and visualisation. Development is back on the evogytis/baltic branch (i.e. here)*. (Github).
28. *beast-mcmc: Bayesian Evolutionary Analysis Sampling Trees*. (Github).
29. *beast-mcmc: Bayesian Evolutionary Analysis Sampling Trees*. (Github).
30. United Nations Statistics Division. Methodology. https://unstats.un.org/unsd/methodology/m49/overview/.