

Emergence of a SARS-CoV-2 E484K variant of interest in Arizona

Peter T. Skidmore¹, Emily A. Kaelin^{1,2}, LaRinda A. Holland¹, Rabia Maqsood¹, Lily I. Wu¹,
Nicholas J. Mellor³, Joy M. Blain³, Valerie Harris⁴, Joshua LaBaer⁴, Vel Murugan⁴,
Efrem S. Lim^{1,2,#}

¹Center for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State University, Tempe, Arizona, USA.

²School of Life Sciences, Arizona State University, Tempe, Arizona, USA.

³OKED Biosciences Genomics Core Facility, Biodesign Institute, Arizona State University, Tempe, Arizona, USA.

⁴Virginia G. Piper Center for Personalized Diagnostics, and ASU Biodesign Clinical testing Laboratory, Biodesign Institute, Arizona State University, Tempe, Arizona, USA.

#Address correspondence to Efrem S. Lim, Efrem.Lim@asu.edu.

Abstract

SARS-CoV-2 is locked in a high-stakes arms race between the dynamics of rising population immunity and escape mutations. The E484K mutation in the spike protein reduces neutralization by post-vaccination sera and monoclonal antibody therapeutics. We detected the emergence of an E484K harboring variant B.1.243.1 from a common circulating variant (B.1.243) in the United States. In contrast to other instances when the E484K mutation was acquired independently in the parental lineage, genomic surveillance indicates that the B.1.243.1 variant of interest is in the process of being established in Arizona and beginning to cross state borders to New Mexico and Texas. Genomic, epidemiologic and phylogenetic evidence indicates that the B.1.243.1 variant of interest is poised to emerge. These findings demonstrate the critical need to continue tracking SARS-CoV-2 in real-time to inform public health strategies, diagnostics, medical countermeasures and vaccines.

Main

The development of multiple effective vaccines is a major advancement in the fight against COVID-19. However, SARS-CoV-2 continues to evolve mutations in its genome as it spreads around the world. Although many mutations have little or no consequence on virus fitness, some mutations affect receptor binding, reduce antibody neutralization or affect diagnostic tests¹⁻⁴. Other mutations have been associated with increased transmission and clinical disease severity^{5,6}. Based on the combination of mutations and associated attributes, variants can be elevated to the classification level of variant of interest (VOI), variant of concern (VOC) and variant of high consequence (VOHC)⁷. Current VOCs include B.1.1.7 that emerged from the UK, B.1.351 that was first identified in South Africa⁸, P.1 that was first described in Brazil⁹, and B.1.427 and B.1.429 that were first detected in California in the US¹⁰. Additionally,

there are 3 VOIs – all of which harbor the E484K mutation in the spike glycoprotein S gene: B.1.525 and B.1.526 both first detected in New York in the US¹¹, and P.2 that was first identified in Brazil¹².

In an effort to provide state-wide genomic surveillance, we sequenced the SARS-CoV-2 genome from 688 positive samples collected from December 28 2020 to March 13 2021 in Arizona, USA. 638 high-quality complete genomes (92.7%) were successfully sequenced that included variants such as B.1.1.7 (n=49, 7.7%), B.1.427/429 (n=214, 33.5%) and P.2 (n=6, 0.9%). We detected 7 genomes associated with a common B.1.243 variant that had acquired an E484K mutation in the spike protein. The novel variant had 11 lineage-defining mutations including V213G and E484K in the spike gene, a 9-nt deletion in ORF1ab (Δ SGF3675-77), a 3-nt insertion in the non-coding intergenic region upstream of the N gene and other synonymous substitutions (**Figure 1A, Supplementary Table 1**). These 11 conserved mutations are distinct from the mutations associated with the parental lineage, B.1.243. The parental B.1.243 lineage is a common circulating variant in the US that was first observed at the start of the pandemic as early as March 2020 (**Figure 1B**, 96.9%). The B.1.243 parent lineage encodes the spike gene D614G substitution, but none of the other concerning mutations (**Figure 1A, Supplementary Table 2**). Therefore, we designate the new E484K harboring variant the provisional name of B.1.243.1.

We next examined the GISAID repository for additional B.1.243.1 genomes to understand its prevalence and geographic distribution. We found that B.1.243.1 is predominantly established in Arizona. Of the 17 cases of B.1.243.1 sequenced to date, 15 cases were from Arizona ranging from February 1 to the most recent case on March 2 (**Figure 1C, Supplementary Table 3**). 1 case was sequenced from a sample collected in Houston, Texas on February 24 and the most recent case from a sample collected in New Mexico on March 8, suggesting that B.1.243.1 has spread to other states in the US. We also identified 2

instances where the parental B.1.243 lineage independently acquired the E484K mutation. However, in contrast to the new B.1.243.1 variant, both genomes lacked the 11 B.1.243.1 lineage-defining mutations and appear to be dead-end transmission events. Phylogenetic analyses support that the B.1.243.1 sequences form a monophyletic clade within the 20A/B.1.243 clade (**Figure 2**). Multiple internal branching observed in the B.1.243.1 clade indicate continued diversification of the lineage sequences. This suggests that B.1.243.1 is being established in circulation within Arizona. In contrast, the two B.1.243 cases bearing the E484K mutation alone were phylogenetically distinct from the B.1.243.1 clade, supporting that they had evolved independently.

Mutations in the SARS-CoV-2 spike protein receptor binding domain (RBD) can affect ACE2 receptor binding and antibody neutralization^{3,13,14}. E484K variants have reduced neutralization by monoclonal antibodies and post-vaccination sera^{1,13,15-17}, potentially complicating antibody-based countermeasures and vaccines. E484K variants have also been identified in reinfection cases, suggesting a role in breakthrough infections^{12,18}. The impact of E484K mutation on spike-ACE2 interaction and binding affinity is less clear¹⁹⁻²¹. Little is known about the functional consequences of the other B.1.243.1 lineage-defining mutations and whether they contribute to the successful transmission in Arizona. Human behavioral risk factors and public policies (e.g. reopening) can also impact the transmissibility of SARS-CoV-2^{22,23}. Due to limited sequencing efforts in the US, the true prevalence and geographic distribution of B.1.243.1 is not well understood. However, this apparent limitation can be resolved with increased sequencing surveillance together with contact tracing. Another limitation of this study was that only 92.7% of samples were successfully sequenced. Most of the partial genomes sequenced were from samples that had lower viral load (high CT values in the 30s), which further reduced the number of sequences that could be used to analyze for this variant.

Based on the mutation profile, regional introduction and phylogenetic evidence, we recommend vigilant monitoring of B.1.243.1 as a potential variant of interest. This study demonstrates the need for sustained genomic surveillance in public health strategies and responses.

Materials and Methods

Study population. As part of SARS-CoV-2 genomic surveillance efforts, saliva samples submitted for COVID-19 testing to the Arizona State University's Biodesign Clinical Testing Laboratory (ABCTL) that tested positive (TaqPath COVID-19 Combo Kit, Applied Biosystems) were randomly selected for next-generation sequencing²⁴. Samples covered a broad distribution from counties across Arizona, United States. This study was approved by the Arizona State University Institutional Review Board.

SARS-CoV-2 sequencing. RNA was extracted from 250 µl of saliva sample (KingFisher Flex, Thermo Scientific) according to the manufacturer's guidelines. First-strand cDNA synthesis was performed using random hexamers (SuperScript III/IV Reverse Transcriptase, Life Technologies), followed by PCR amplification of tiled amplicons spanning the SARS-CoV-2 genome (Swift Normalase Amplicon Panel, Swift Biosciences) and library construction. Libraries were sequenced on the Illumina MiSeq (version 2, 2x150) and NextSeq 500 (version 2.5, 2x150, mid or high output).

Sequencing analysis. Illumina sequencing reads were quality filtered to remove adaptors and low-quality bases using BBTools. High-quality-filtered reads were mapped to the SARS-CoV-2 Wuhan1 reference genome (NC_045512.2) using BWA-MEM²⁵ and amplicon primers were trimmed using Primerclip (version 0.3.8)²⁶. Consensus sequence was called using iVar (version 1.0; parameters -q 20, -t 0.75, -m 20, -n N)²⁷. Sequence alignments were performed with

MAFFT (version 7.471)²⁸ and variant calling using Geneious Prime (version 2021). Sequences used in phylogenetic analysis include the global sequences from GISAID²⁹ subset to 500 sequences, a subset of 100 B.1.243 sequences from all B.1.243 GISAID sequences, and the 17 B.1.243.1 lineage sequences. Phylogenetic reconstruction was performed with IQTree (version 2.0.3³⁰, `iqtree -nt AUTO -bb 1000 -m MFP -mset GTR`), and Augur (version 11.3.0)³¹.

Data availability. Sequence data have been deposited in GISAID.

Competing interests

The authors declare no competing interests.

Acknowledgements

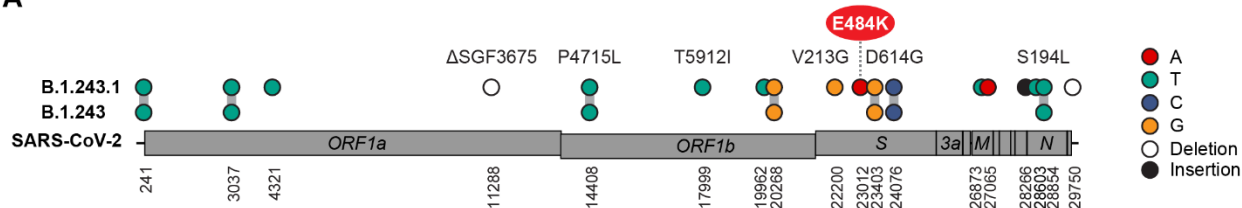
We gratefully acknowledge the authors from originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID initiative, on which part of the research is based. This work was supported in part by the Arizona State University Knowledge Enterprise and Arizona Department of Health Services.

Contributions

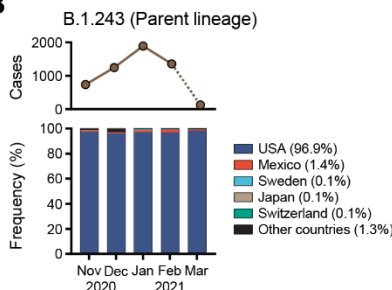
Methodology: P.T.S., E.A.K., L.A.H., R.M., L.I.W., N.J.M., J.M.B., V.H., E.S.L.; Investigation: P.T.S., E.A.K., R.M., E.S.L.; Resources: L.I.W., V.H., J.L., V.M.; Data curation: P.T.S., E.A.K., R.M.; Writing-original draft: P.T.S., E.A.K., R.M., E.S.L.; Writing-review and editing: P.T.S., E.A.K., L.A.H., R.M., E.S.L.; Supervision: J.L., V.M., E.S.L.; Conceptualization: E.S.L.; Funding acquisition: E.S.L. All authors reviewed and approved the final manuscript.

Figures

A



B



C

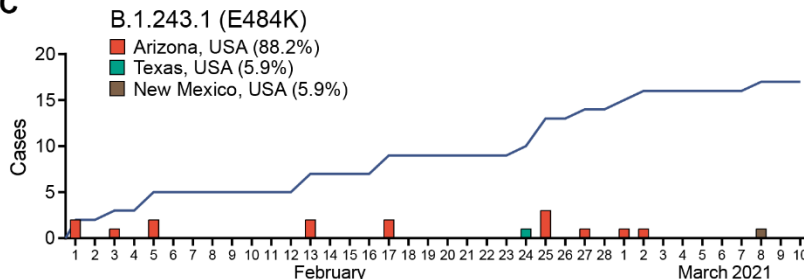


Figure 1. Emergence of E484K harboring B.1.243.1 variant in Arizona, USA. (A) B.1.243.1 lineage-defining mutations are shown on the SARS-CoV-2 genome. Mutations are shown in reference to the SARS-CoV-2 Wuhan-1 genome position (NC_045512.2). **(B)** Global prevalence of B.1.243 parental lineage from November 2020 to March 2021. Monthly number of B.1.243 cases (top) and distribution by country (bottom) are shown. Average frequency for each country is shown in parenthesis. Sequences from March are under-reported at the time of this reporting (indicated by dashed line). **(C)** B.1.243.1 case incidence reporting from February to March 2021. Cumulative case incidence is plotted as a line graph.

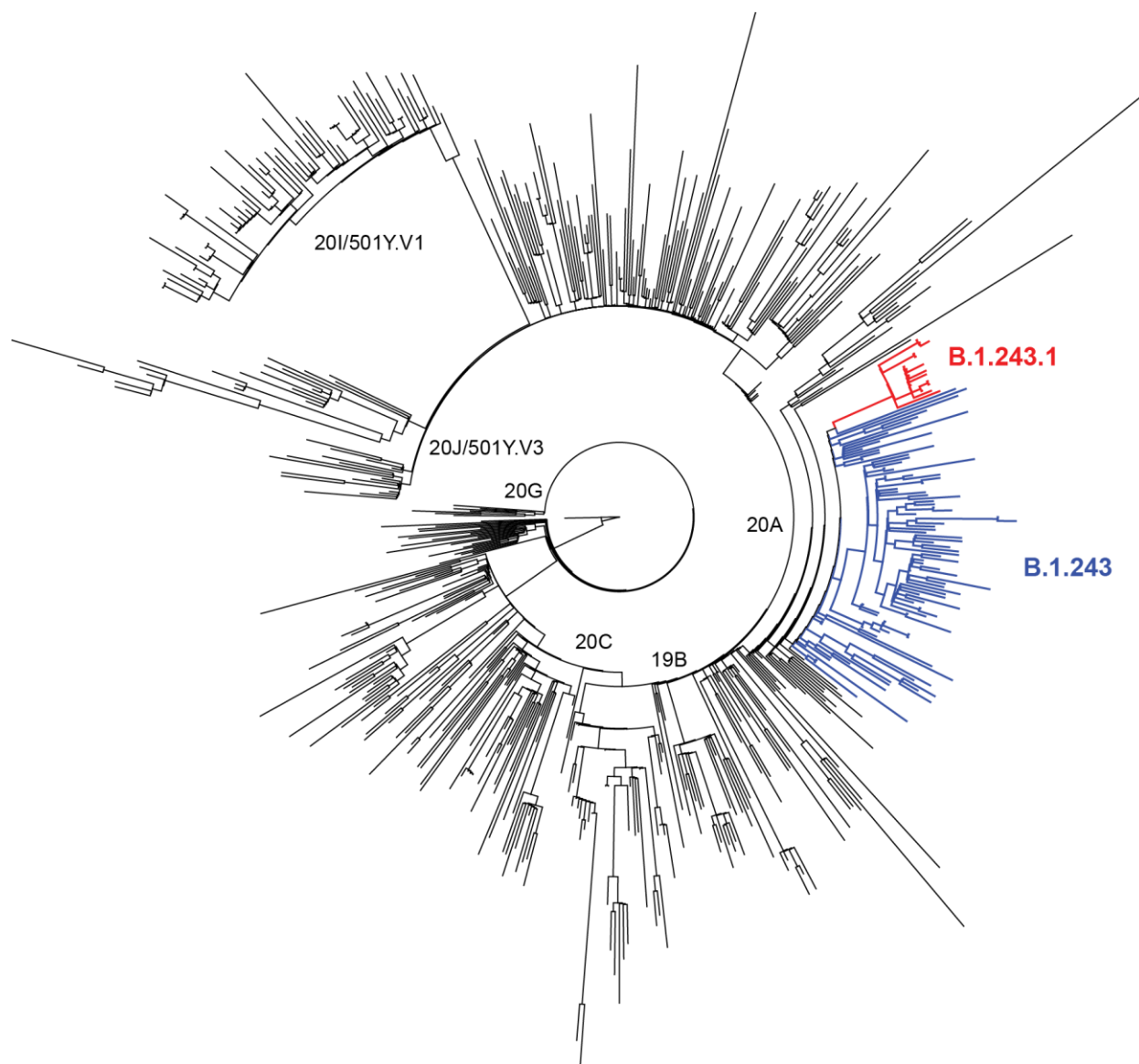


Figure 2. Phylogenetic relationship of the novel B.1.243.1 lineage. Maximum likelihood phylogeny of diverse SARS-CoV-2 sequences including 500 representative global sequences, 100 B.1.243 parent lineage sequences, and the 17 B.1.243.1 sequences. The novel B.1.243.1 lineage is indicated in red branches (clade bootstrap support: 100), and the parental B.1.241 lineage in blue.

Supplementary Table 1. B.1.243.1 lineage-defining mutations.

Gene	Nucleotide change	Amino acid change	Number of B.1.243.1 genomes with mutation
ORF1ab	C4321T	Synonymous	17/17
ORF1ab	11288-11296 deletion	SGF 3675-3677 deletion	17/17
ORF1ab	C17999T	T5912I	17/17
ORF1ab	G19962T	Synonymous	17/17
S	T22200G	V213G	17/17
S	G23012A	E484K	17/17
M	C26873T	Synonymous	17/17
M	G27065A	Synonymous	17/17
Non-coding, upstream of N	28266 GCC insertion	Non-coding	17/17
N	C28603T	Synonymous	17/17
3' UTR	29750-29761 deletion	Non-coding	13/13*

All genome positions are in reference to the SARS-CoV-2 Wuhan-1 sequence (NC_045512.2).

*4 sequences were omitted due to ambiguous (N) nucleotides.

Supplementary Table 2. Lineage-defining mutations in parental B.1.243 lineage.

Gene	Nucleotide change	Amino acid change	Percent of B.1.243 genomes with mutation
5' UTR	C241T	Non-coding	98.0%
ORF1ab	C3037T	Synonymous	99.4%
ORF1ab	C14408T	P4715L	98.6%
ORF1ab	A20268G	Synonymous	95.5%
S	A23403G	D614G	100%
S	T24076C	Synonymous	98.9%
N	C28854T	S194L	97.9%

Based on 7,211 global B.1.243 genomes downloaded from GISAID on 3/20/21.

Supplementary Table 3: B.1.243.1 sequences.

Name	GISAID Accession	Location	Collection Date
hCoV-19/USA/AZ-ASU2621/2021	EPI_ISL_1364812	Arizona	2/1/2021
hCoV-19/USA/AZ-ASU2625/2021	EPI_ISL_1364644	Arizona	2/1/2021
hCoV-19/USA/AZ-ASU2857/2021	EPI_ISL_1364649	Arizona	2/3/2021
hCoV-19/USA/AZ-CDC-21801839/2021	EPI_ISL_1090700	Arizona	2/5/2021
hCoV-19/USA/AZ-CDC-21802041/2021	EPI_ISL_1090853	Arizona	2/5/2021
hCoV-19/USA/AZ-CDC-22062741/2021	EPI_ISL_1139766	Arizona	2/13/2021
hCoV-19/USA/AZ-ASU2754/2021	EPI_ISL_1364775	Arizona	2/13/2021
hCoV-19/USA/AZ-ASU3132/2021	EPI_ISL_1365543	Arizona	2/17/2021
hCoV-19/USA/AZ-ASU2925/2021	EPI_ISL_1365483	Arizona	2/17/2021
hCoV-19/USA/TX-HMH-MCoV-29140/2021	EPI_ISL_1303700	Texas	2/24/2021
hCoV-19/USA/AZ-ASU2540/2021	EPI_ISL_1291671	Arizona	2/25/2021
hCoV-19/USA/AZ-TG758899/2021	EPI_ISL_1292269	Arizona	2/25/2021
hCoV-19/USA/AZ-TG758666/2021	EPI_ISL_1292117	Arizona	2/25/2021
hCoV-19/USA/AZ-CDC-22555310/2021	EPI_ISL_1290985	Arizona	2/27/2021
hCoV-19/USA/AZ-CDC-22554229/2021	EPI_ISL_1290992	Arizona	3/1/2021
hCoV-19/USA/AZ-TG761699/2021	EPI_ISL_1296905	Arizona	3/2/2021
hCoV-19/USA/NMDOH-2021075279/2021	EPI_ISL_1340909	New Mexico	3/8/2021

References

1. Wu, K., *et al.* mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. *bioRxiv* (2021).
2. Xie, X., *et al.* Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat Med* (2021).
3. Starr, T.N., *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310 e1220 (2020).
4. Kemp, S.A., *et al.* Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70. *bioRxiv* (2021).
5. Liu, Y., *et al.* The N501Y spike substitution enhances SARS-CoV-2 transmission. *bioRxiv* (2021).
6. Challen, R., *et al.* Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* **372**, n579 (2021).
7. CDC. SARS-CoV-2 Variants. (2021).
8. Tegally, H., *et al.* Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature* (2021).
9. Faria, N.R., *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological* (2021).
10. Peng, J., *et al.* Estimation of secondary household attack rates for emergent SARS-CoV-2 variants detected by genomic surveillance at a community-based testing site in San Francisco. *medRxiv* (2021).
11. Annavajhala, M.K., *et al.* A Novel SARS-CoV-2 Variant of Concern, B.1.526, Identified in New York. *medRxiv* (2021).
12. Nonaka, C.K.V., *et al.* Genomic Evidence of SARS-CoV-2 Reinfection Involving E484K Spike Mutation, Brazil. *Emerg Infect Dis* **27**(2021).
13. Liu, Z., *et al.* Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host & Microbe* **29**, 477-488.e474 (2021).
14. Greaney, A.J., *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463-476 e466 (2021).
15. Garcia-Beltran, W.F., *et al.* Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* (2021).
16. Li, Q., *et al.* SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape. *Cell* (2021).
17. Chen, R.E., *et al.* Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat Med* (2021).
18. Naveca, F., *et al.* SARS-CoV-2 reinfection by the new Variant of Concern (VOC) P.1 in Amazonas, Brazil. *Virological.org* (2021).
19. Wang, W.B., *et al.* E484K mutation in SARS-CoV-2 RBD enhances binding affinity with hACE2 but reduces interactions with neutralizing antibodies and nanobodies: Binding free energy calculation studies. *bioRxiv*, 2021.2002.2017.431566 (2021).
20. Villoutreix, B.O., Calvez, V., Marcelin, A.-G. & Khatib, A.-M. In Silico Investigation of the New UK (B.1.1.7) and South African (501Y.V2) SARS-CoV-2 Variants with a Focus at the ACE2–Spike RBD Interface. *International Journal of Molecular Sciences* **22**, 1695 (2021).
21. Cheng, M.H., Krieger, J.M., Kaynak, B., Arditi, M. & Bahar, I. Impact of South African 501.V2 Variant on SARS-CoV-2 Spike Infectivity and Neutralization: A Structure-based Computational Assessment. *bioRxiv*, 2021.2001.2010.426143 (2021).

22. Adam, D.C., *et al.* Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med* **26**, 1714-1719 (2020).
23. Brauner, J.M., *et al.* Inferring the effectiveness of government interventions against COVID-19. *Science* **371**(2021).
24. Holland, L.A., *et al.* An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *J Virol* **94**(2020).
25. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **1303.3997**(2013).
26. Swift. primerclip. (2021).
27. 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).
28. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).
29. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33-46 (2017).
30. 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).
31. Huddleston, J., *et al.* Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *Journal of Open Source Software* (2021).