

From elimination to suppression: genomic epidemiology of a large Delta SARS-CoV-2 outbreak in Aotearoa New Zealand

Lauren Jelley^{*1,2}, Jordan Douglas^{*3}, Xiaoyun Ren¹, David Winter¹, Andrea McNeill¹, Sue Huang¹, Nigel French⁴, David Welch², James Hadfield⁵, Joep de Ligt¹, Jemma L Geoghegan^{1,2,#}

¹Institute of Environmental Science and Research, Wellington, New Zealand.

²Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand.

³Centre for Computational Evolution, School of Computer Science, University of Auckland, Auckland, New Zealand.

⁴Tāwharau Ora/School of Veterinary Science, Massey University, Palmerston North, New Zealand.

⁵Fred Hutchinson Cancer Research Centre, Seattle, Washington, USA.

*contributed equally

#author for correspondence: jemma.geoghegan@otago.ac.nz

Keywords: SARS-CoV-2, genomic epidemiology, phylogenetics, phylodynamics, epidemiology, Delta variant, virus evolution, ORF7a, Variant of Concern, COVID-19.

Abstract.

New Zealand's COVID-19 elimination strategy heavily relied on the use of genomics to inform contact tracing, linking cases to the border and to clusters during community outbreaks. In August 2021, New Zealand entered its second nationwide lockdown after the detection of a single community case with no immediately apparent epidemiological link to the border. This incursion resulted in the largest outbreak seen in New Zealand caused by the Delta Variant of Concern. Here we generated 3806 high quality SARS-CoV-2 genomes from cases reported in New Zealand between 17 August and 1 December 2021, representing 43% of reported cases. We detected wide geographical spread coupled with undetected community transmission, characterised by the apparent extinction and reappearance of genomically linked clusters. We also identified the emergence, and near replacement, of genomes possessing a 10-nucleotide frameshift deletion that caused the likely truncation of accessory protein ORF7a. By early October, New Zealand moved from elimination to suppression and the role of genomics changed markedly from being used to track and trace, towards population-level surveillance.

Introduction.

Since early 2020, the World Health Organisation has monitored the global evolution of *Severe acute respiratory coronavirus 2* (SARS-CoV-2). As new genomic variants emerged that posed an increased threat to public health, they have been classified as 'Variants of Concern' (VOC), 'Variants of Interest' (VOI) or 'Variants Under Monitoring' (VUM), and most VOCs and VOIs are designated a Greek letter. Such variants typically have one or more of an increase in transmissibility, a change in virulence, and ability to escape pre-existing immunity compared to other genomic variants.

In December 2020, three distinct but related lineages of SARS-CoV-2 were detected in India, named subvariants B.1.617.1, B.1.671.2 and B.1.617.3¹. From these lineages, B.1.617.2 outcompeted the others to become dominant in India and was soon after characterised as a VOC, denoted Delta. By March 2021, India experienced a significant second wave of the pandemic associated with the Delta variant. Just two months later, reported cases in India accounted for half of all global cases and quickly overwhelmed healthcare services². The Delta variant rapidly spread around the world and became dominant globally during 2021, largely displacing Alpha (B.1.1.7) and other variants.

The Delta variant possesses at least 13 non-synonymous mutations compared to ancestral variants and its growth advantage can be explained primarily due to both immune evasion and a 40-60% increase in transmissibility compared to Alpha^{3,4}. The astonishing speed at which Delta spread led to outbreaks in many countries that had previously achieved elimination. For example,

Delta outbreaks in Australia, Fiji, Vietnam, Singapore and Taiwan demonstrated how previous public health measures used to control COVID-19 were less effective against this highly transmissible variant⁵⁻⁹.

New Zealand eliminated community transmission of SARS-CoV-2 by mid 2020 and continued to pursue a zero-COVID policy up until late in 2021¹⁰. Several small but quickly controlled SARS-CoV-2 outbreaks were detected in the community since the virus was first eliminated¹¹. But for the most part, cases of COVID-19 were largely restricted to managed isolation and quarantine facilities at the border, where returning New Zealanders were required to undergo at least 14 days of quarantine. These border control measures averted numerous virus incursions and, as a result, New Zealand saw an increase in life expectancy over the first two years of the pandemic^{12,13}.

In August 2021, following a brief period of reciprocal quarantine-free travel with Australia, New Zealand entered its second nationwide lockdown after the detection of a single COVID-19 community case in Auckland with no immediately apparent epidemiological link to the border. Between 17 August and 1 December 2021 there were 8974 cases of COVID-19 reported in the community. In early October, due to the inability to eliminate Delta even under the most stringent lockdown measures, and with the aid of high vaccination rates, New Zealand moved from elimination to suppression, aimed at minimisation and protection^{14,15}. Stay-at-home orders persisted in Auckland and sporadically across various regions of the country until early December. By the end of 2021 this Delta outbreak was the largest single-origin outbreak in New Zealand - the second largest occurred in August 2020 with fewer than 200 reported community cases¹⁶.

Here we describe a large-scale outbreak of SARS-CoV-2 in New Zealand and show how genomics was used in real-time to first help track and trace cases of COVID-19 in the community and, later, with the change to suppression, to monitor its spread and evolution. This genomic surveillance identified numerous examples of cryptic virus transmission by the apparent extinction and reappearance of genomically linked clusters. Additionally, we identified the emergence and dominance of a lineage possessing a 10-nucleotide frameshift deletion likely rendering an accessory protein, encoded by the ORF7a gene, functionally impaired, within this outbreak.

Methods.

Ethics statement

Nasopharyngeal samples that had positive results for SARS-CoV-2 by real-time reverse transcription PCR were obtained from medical diagnostic laboratories located throughout New Zealand. Under contract for the New Zealand Ministry of Health, the Institute of Environmental

Science and Research has approval to conduct genomic sequencing and phylogenetic analysis for surveillance of notifiable diseases.

Genomic sequencing of SARS-CoV-2

For cases reported between 17 August and 1 December 2021 a random proportion of COVID-19 community cases were referred to the Institute of Environmental Science and Research, New Zealand for genome sequencing. In brief, viral extracts were prepared from respiratory tract samples in which SARS-CoV-2 was detected by rRT-PCR. Extracted RNA was subjected to whole-genome sequencing using the Oxford Nanopore Technologies R9.4 chemistry by following the Midnight protocol v6, designed by Freed et al.¹⁷, which contains a 1200-bp primer set tiling the SARS-CoV-2 genome. Consensus genomes were generated through a standardised pipeline (https://github.com/ESR-NZ/NZ_SARS-CoV-2_genomics) based on the original ARTIC bioinformatics pipeline (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>).

Phylogenetic analysis

A total of 3806 high quality genomes were generated from New Zealand's 2021 Delta outbreak sampled between 17 August and 1 December. Genomes were assigned to the B.1.617.2 lineage using PANGO (v.3.1.20)¹ (<https://github.com/cov-lineages/pangolin/releases/tag/v3.1.20>). Four genomes sequenced from New Zealand's managed isolation and quarantine facility were also included that were genetically identical to the consensus sequences of the first reported community case and were sampled the week prior to the outbreak. These genomes were aligned along with 1500 global genomes uniformly selected at random from GISAID¹⁸ sampled over the same time, including four genomes sequenced in New South Wales, Australia that were genetically identical to the presumed index cases in New Zealand and were sampled 10 days prior to the first reported community case (GISAID accessions: EPI_ISL_3426406, EPI_ISL_3426050, EPI_ISL_3426425 and EPI_ISL_3426179).

Genomes were aligned using Nextalign¹⁹ with Wuhan-Hu-1 (NC_045512.2) used as a reference. A maximum likelihood phylogenetic tree was estimated using IQ-TREE (v 1.6.8)²⁰ using the Hasegawa-Kishino-Yano (HKY+ Γ)²¹ nucleotide substitution model with a gamma distributed rate variation among sites (the best fit model was determined by ModelFinder²²), and branch support was assessed using the ultrafast bootstrap method²³. The resulting phylogenetic tree was visualised and displayed using GrapeTree²⁴ and annotated by the location (i.e. District Health Board) from where genomes were sampled.

ORF7a deletion and protein characterisation

Genomes sampled between 17 August and 1 December 2021 were annotated based on whether they possessed a 10-nucleotide deletion within the Open Reading Frame (ORF) 7a between

positions 27694 and 27704 (see Supplementary Figure 1). This resulted in a total of 2309 and 1126 genomes with and without the ORF7a deletion, respectively. The remaining 371 consensus genomes possessed ambiguous nucleotides within this region and were therefore excluded from further analysis. The transmembrane properties of ORF7a were modelled and annotated using Protter²⁵.

Estimating the effective reproduction number

To estimate the effective reproduction number (R_{eff}) through time for genomes with and without the ORF7a deletion, a Bayesian birth-death skyline model²⁶ was implemented in BEAST 2.5²⁷. First, genomes were aligned as above and subsampled down to 33% ($n=374$ for wildtype and $n=770$ for truncated) to facilitate convergence of the Markov chain Monte Carlo (MCMC) algorithm. R_{eff} was estimated for the period between the root of the tree and detection of the first community case, and then every 10-day period up until 1 December 2021 (but due to high uncertainty in R_{eff} approaching the most recent sample, we only report these estimates up until 1 November). Consecutive estimates were smoothed using an Ornstein-Uhlenbeck smoothing prior. Nucleotide evolution was modelled using the HKY substitution model with frequencies estimated. MCMC chains were run until the effective sample size of reported parameters exceeded 200. To help speed up MCMC convergence, we used the efficient tree operators in the BICEPS package²⁷ and adaptable operator weighting²⁸.

Results.

We generated 3806 high quality genomes, designated as Delta VOC (B.1.617.2; lineage AY.39.1.1), sampled between 17 August and 1 December 2021. To inform the goal of elimination of COVID-19 in the community, between 17 August and 3 October, >88% of reported cases resulted in high-quality genomes, which directly aided contact tracing efforts (Figure 1). The proportion of cases subject to genome sequencing fell to an average of 33% from 4 October to 1 December as a consequence of the change in strategy from elimination to suppression and the increase in reported cases. The number of reported cases during the Delta outbreak peaked on 16 November and case numbers have since declined markedly. As of March 2022, Delta cases continue to be detected through routine genomic surveillance, although remaining at low levels among a background of cases with the highly transmissible Omicron variant²⁹. The data generated originated from 15 out of 20 District Health Boards located across the country (Figure 1). In line with the geographic location of the majority of cases, the vast majority of genomes sampled over this time frame were from Auckland (19%), Counties Manukau (38%) and Waitemata (31%) - the three metropolitan District Health Boards that surround Auckland, New Zealand's most populous city (Figure 1).

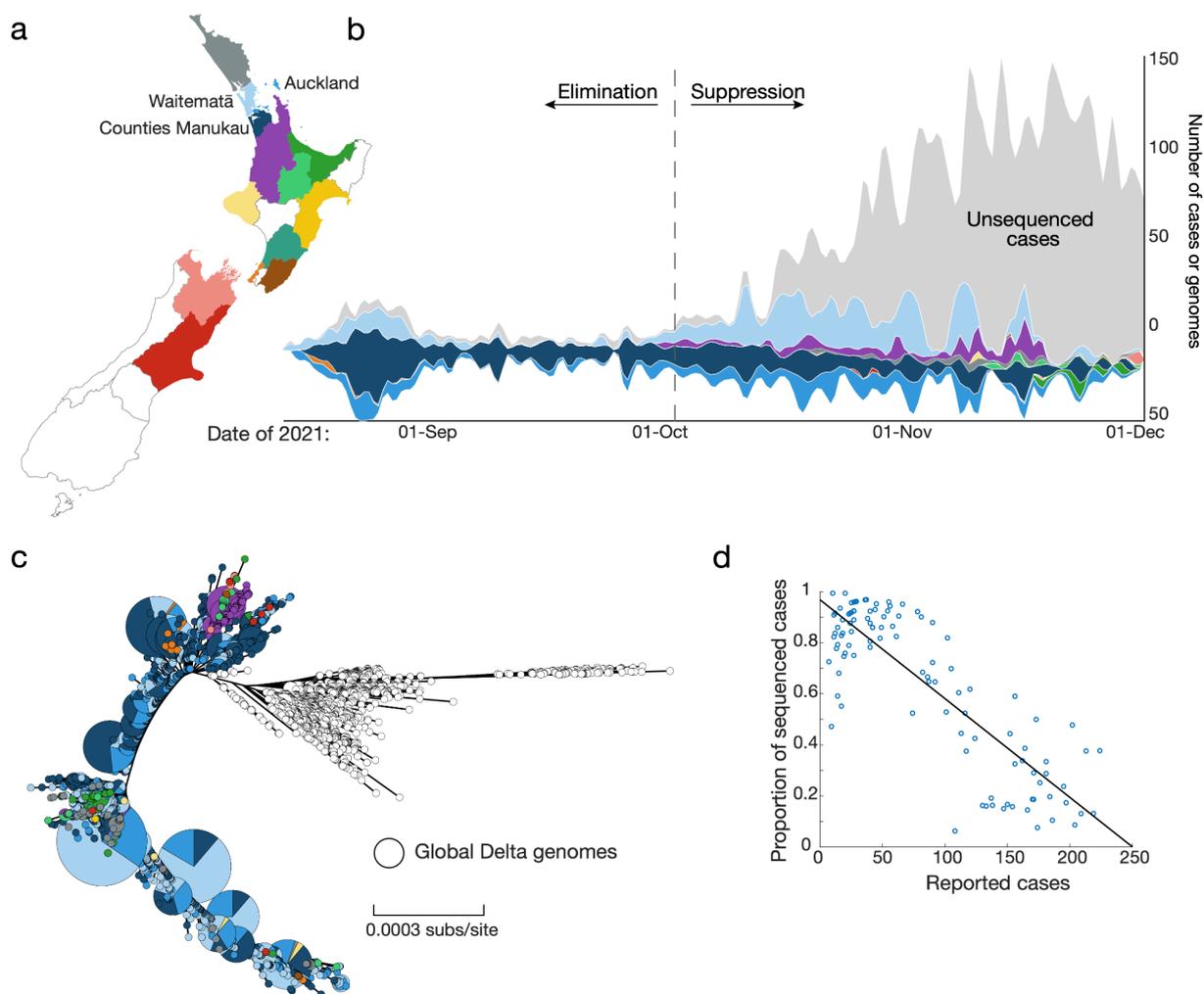


Figure 1. (a) Map of New Zealand where regions are categorised by District Health Board. Coloured regions indicate positive COVID-19 cases were reported in that location between 17 August and 1 December. (b) Streamgraph illustrating the proportion of genomes sequenced coloured by region, corresponding to the map in (a). (c) Maximum likelihood phylogenetic tree displayed as a GrapeTree²⁴ coloured by region along with global genomes (white). Branch lengths are scaled by the number of nucleotide substitutions per site. (d) Proportion of reported cases sequenced versus number of reported cases with least-squares line.

New Zealand's 2021 Delta outbreak was genomically linked to a single introduction into the community (Figure 1). While no definitive epidemiological link to the border could be established, genome sequencing revealed four genomes sampled from a managed isolation and quarantine facility in Auckland that were genetically indistinguishable to the consensus genome of the first reported case in the community. These cases originated from two separate travel groups who had recently returned from New South Wales, Australia in the week prior to the community outbreak. In

addition, several genetically indistinguishable genomes were sampled from the New South Wales community during this time.

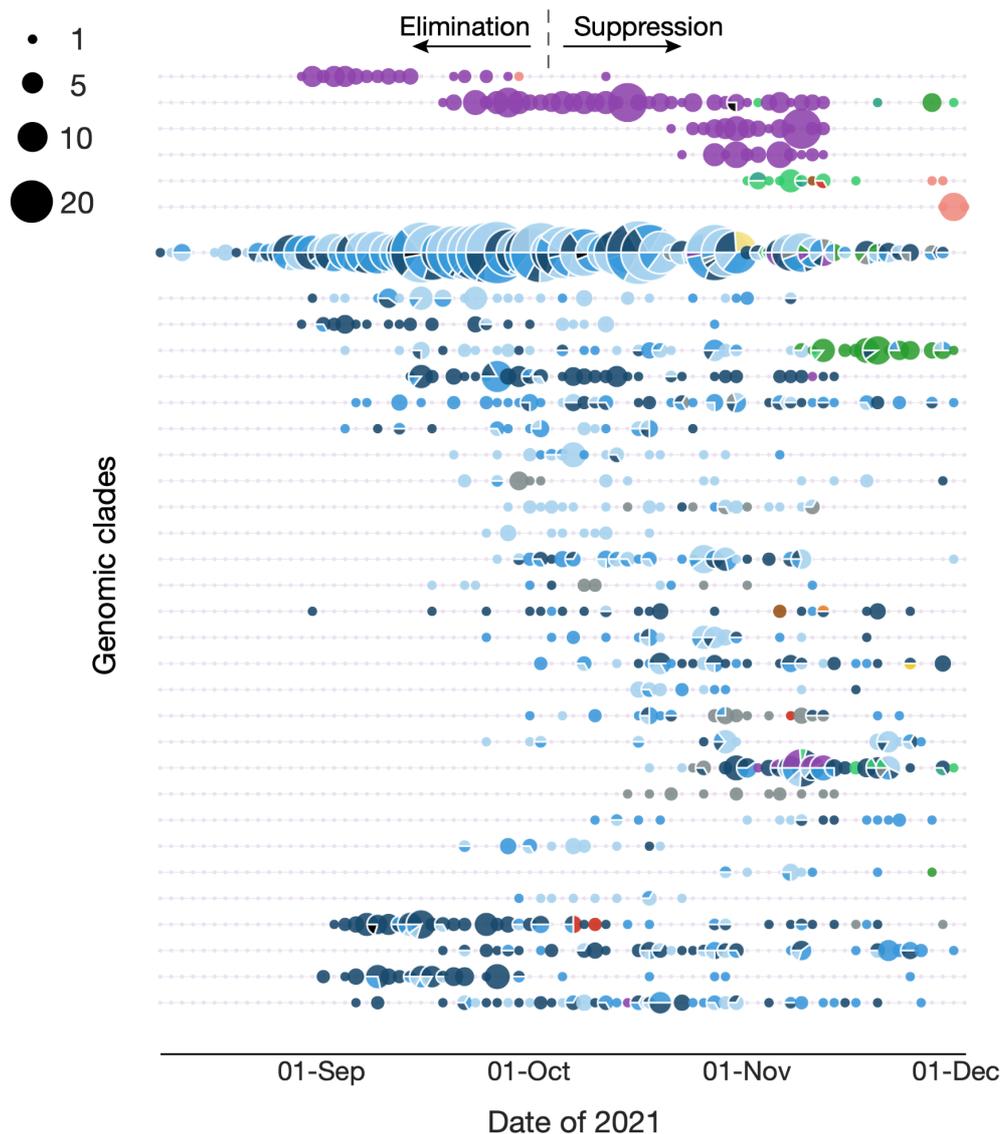


Figure 2. Bead plot showing a summary of active genomic clades per day between 17 August and 1 December 2021 coloured by the geographical region (District Health Board) shown in Figure 1a. Each string of beads represents a genomic clade and the size of each bead corresponds to the number of reported cases per day within each clade. Genomic clades were defined as involving at least one single nucleotide polymorphism (SNP) and including at least 10 descendants.

Until early October 2021, genomics was used to trace every reported case of COVID-19 in New Zealand. This was hampered by known genomic clades going unsampled before reappearing weeks later, suggesting substantial undetected community transmission. As case numbers rose and New Zealand moved to a suppression strategy the focus moved to monitoring active genomic

clades (Figure 2). Cases reported from Auckland's metropolitan areas were evident in nearly all clades whereas cases reported from more regional localities showed much less genomic variation. There was strong evidence of transmission 'seeding' from Auckland's metropolitan areas into other regions, resulting in the emergence of new, dominant clades that amplified, diverged and ultimately spread into further locations. An example is the Waikato region (purple in Figure 2), where there was evidence of widespread community transmission of a single, dominant clade that subsequently diverged into four subclades, before a period characterised by cryptic spread and transmission to other regions.

The outbreak was characterised by the emergence of a 10-nucleotide frameshift deletion in the ORF7a gene, which produces one of 11 accessory proteins in the SARS-CoV-2 genome (Figure 3; Supplementary Figure 1). This deletion spanned positions 27694 and 27704 and resulted in a predicted truncated protein of only 103 amino acids (compared to the 121 long amino acid wildtype product without the deletion), containing three altered amino acids prior to a premature stop codon. The first virus genome possessing the ORF7a deletion was sequenced on 27 August and by early October genomes with this deletion accounted for over 90% of sequenced cases (Figure 3). There was little difference in the age distribution in cases with the deletion (mean age 31) and those without (mean age 27) suggesting viruses with the deletion was not more prevalent in any one particular age group (Figure 3).

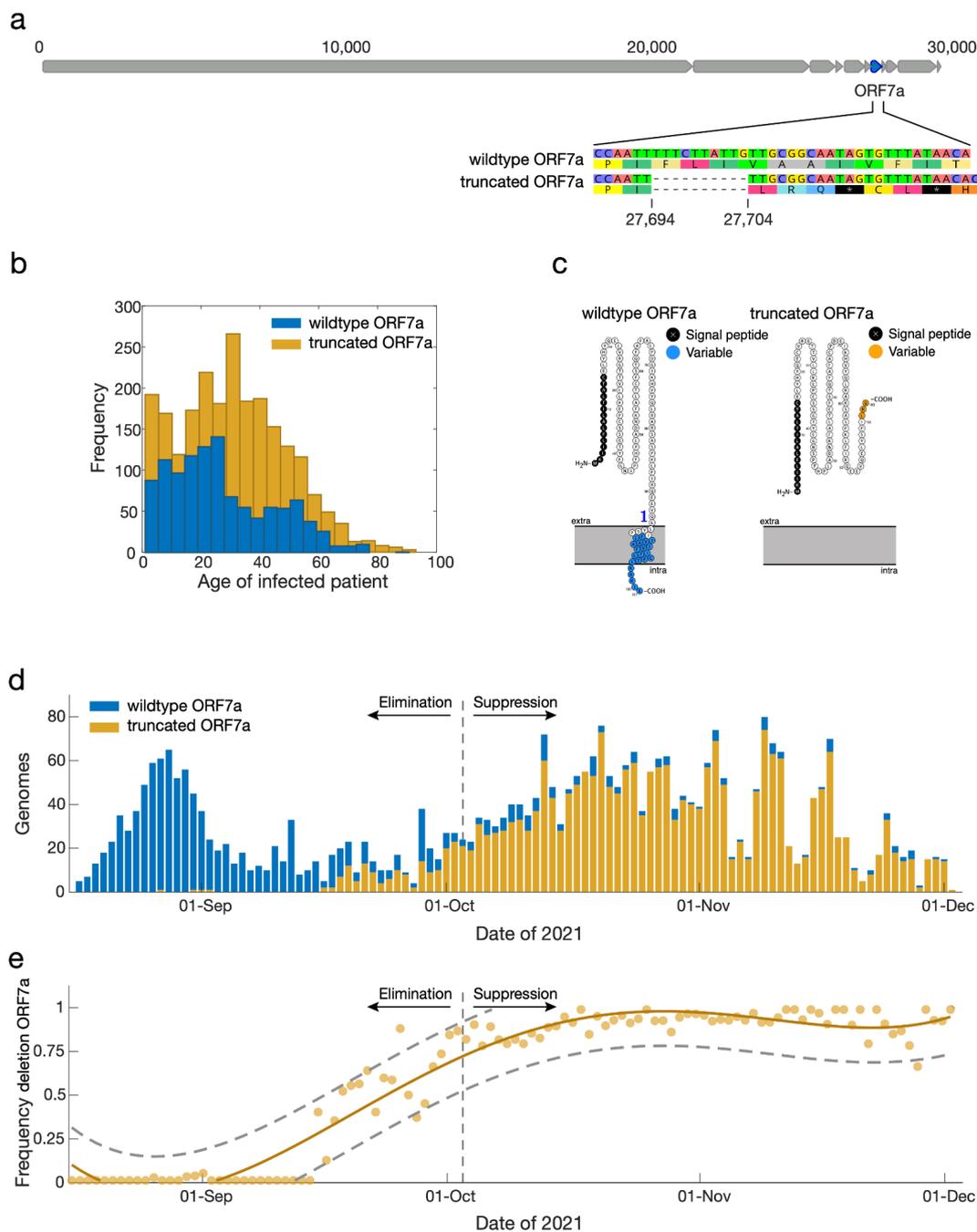


Figure 3. (a) Genome structure of SARS-CoV-2 illustrating the position of ORF7a along with an alignment showing the 10-nucleotide deletion between positions 27694 and 27704 and resulting frameshift mutation. (b) Age distribution of people infected with SARS-CoV-2 possessing virus genomes with (orange) and without (blue) the 10-nucleotide deletion in ORF7a. (c) Protein structure of the truncated and wildtype accessory protein 7a variants predicted by Protter²⁵. A transmembrane helix is shown crossing the endoplasmic reticulum membrane for the wildtype variant and the variable amino acids are indicated. (d) Genomes sequenced through time with (orange) and without (blue) the 10-nucleotide deletion in ORF7a. (e) The frequency of genomes with the truncated ORF7a peptide (orange line), a 95% prediction interval (grey dashed lines) and the raw data (orange points) over time.

The ORF7a deletion is predicted to have truncated three structural features at the C-terminus of accessory protein 7a (Figure 3):

- i) the transmembrane helix;
- ii) the endoplasmic reticulum (ER) localisation signal KRKIE; and
- iii) the post-translational modification site K119³⁰.

This deletion would likely render accessory protein 7a unable to carry out regular functions in the ER/Golgi complex. Despite this defect, viruses with the truncated ORF7a lineage clearly remained viable and indeed quickly became the dominant Delta lineage in New Zealand.

We estimated the R_{eff} of both lineages through time (Figure 4). The wildtype ORF7a lineage peaked shortly after the first cases were discovered, with an R_{eff} of 6.8 (and a 95% credible interval (CI) of 2.1-15). Following this, R_{eff} declined rapidly to around 1 for the remainder of the outbreak, likely as a result of a stringent public health response and high vaccination rates (86.4% of the eligible (>12 years of age) population by 1 December). Similarly, the truncated ORF7a lineage peaked shortly after it was first detected, with a mean R_{eff} of 2.1 (95% CI: 0.1-4.5). Although the truncated ORF7a lineage saw some initial success, its R_{eff} gradually declined to around 1. There is no evidence that the wildtype nor truncated ORF7a lineages have any significant differences in transmissibility.

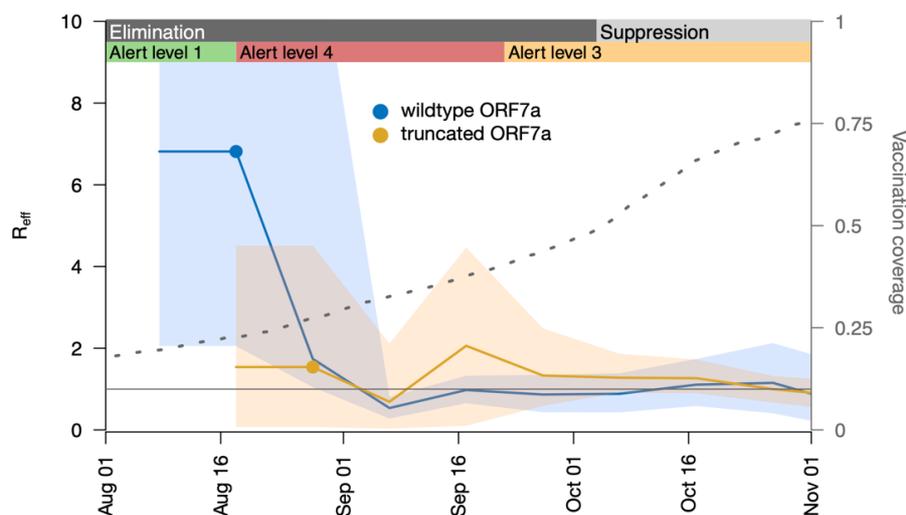


Figure 4. Mean estimate of the effective reproduction number (R_{eff}) over time (solid line) and its 95% credible interval (shaded area), based on genomes with the truncated (orange) or wildtype (blue) ORF7a. One R_{eff} estimate is made per 10-day period following detection of the first community case (filled circle) for either lineage, plus one additional estimate between this period and the first transmission event (i.e. the root of the tree). The right axis shows the proportion of the eligible population (12+ years old) who had received two doses of the Pfizer-BioNTech COVID-19 vaccine. The nationwide change from elimination to suppression strategy is indicated, as are regional changes between mitigation settings in Auckland (the epicentre of the outbreak), with 'Alert level 4' being the most stringent¹⁰.

Discussion.

New Zealand's elimination strategy for controlling SARS-CoV-2 has been hailed a success, and the nation saw low levels of mortality during the first two years of the pandemic^{12,13}. Genomics played an essential role in aiding the COVID-19 elimination strategy, where, until early October 2021, all positive cases were referred for sequencing and integrated with epidemiological data to directly inform control efforts. Genomics of SARS-CoV-2 helped identify cluster membership where epidemiological links were unclear and link community cases to the border^{11,16,31}. Here we show that a single incursion event, possibly from a managed isolation and quarantine facility was responsible for New Zealand's third wave of COVID-19. Although no clear epidemiological link was ever established, New Zealand's August 2021 Delta outbreak was genomically linked to cases at the border as well as to a concurrent outbreak in New South Wales, Australia.

An estimate of the rate of transmission showed that R_{eff} continued to remain ~ 1 despite the most stringent lockdown measures¹⁰. This was in contrast to the effect of lockdown measures on ancestral genomic variants in New Zealand where R_{eff} fell close to zero during the first wave in 2020^{32,33} even following multiple superspreading events, thus enabling virus elimination. During the elimination phase of this Delta outbreak, genomic clades often appeared to go extinct only to reappear weeks later, suggestive of widespread undetected transmission. With the number of positive cases increasing, lockdowns unable to prevent onward transmission, and with increasing vaccination rates, New Zealand moved from an elimination strategy to a suppression strategy by early October^{14,15}. The change in strategy changed the role of genomics, which shifted from one focused on contact tracing and origin detection to one targeted at the level of the New Zealand population, with a specific focus on viral evolution.

Genomic surveillance showed that the New Zealand Delta outbreak was characterised by the emergence and subsequent increase in frequency of genomes containing a 10-nucleotide frameshift deletion in ORF7a. The polymorphic nature of this deletion provided greater resolution to the epidemiological picture, enabling cases to be linked to transmission chains and therefore aiding contact tracing efforts during the elimination period among otherwise relatively clonal genomes. The ORF7a gene itself encodes accessory protein 7a, which has strong sequence similarity with its ortholog within the SARS-CoV-1 genome³⁴. Accessory protein 7a is a transmembrane protein with a seven-stranded β -sandwich fold³⁵. Although considered non-essential, it is involved in a range of functions including antagonisation of the host interferon type 1 response³⁶, binding to CD14+ monocytes in human peripheral blood³⁷, involvement in protein trafficking³⁸ and regulation of the ORF7b peptide³⁹, and is known to induce apoptosis when overexpressed^{40,41}.

Although highly variable in position and length, frameshift deletions that induce truncation of accessory protein 7a are not unique. There are numerous independent occurrences of such a

mutation across both A and B lineages of SARS-CoV-2, including among VOCs⁴²⁻⁴⁹. Global genomic surveillance of SARS-CoV-2 has shown that mutations frequently occur at the downstream regions of ORF7a with such changes often leading to premature stop codons resulting in protein products with reduced functional activity. Such mutations in ORF7a appear to be associated with various viral mechanisms, including reduced viral suppression of the immune response⁵⁰, an increase in viral progeny⁴⁶, and a seemingly rapid increase in viral spread⁴⁴. Indeed, such “dispensable” accessory proteins are known to gain and lose functional activities quite rapidly in evolutionary time, across a wide range of viral taxonomies⁵¹⁻⁵³ – perhaps as a means to better explore their fitness landscapes and co-evolve with their host. This certainly appears to be the case for accessory protein 7a, which has repeatedly seen sweeping structural mutations across a range of geographical regions, while still maintaining a high level of fitness.

The truncated ORF7a lineage saw some initial success in New Zealand and quickly became the dominant variant in the community. Nevertheless, this may have simply been a founder effect that coincided with a relaxation of public health restrictions. Overall, there are too many confounding factors and an insignificant difference in reproduction number to determine whether the truncated variant had any enhanced transmissibility over the wildtype.

Overall, genomics has played a key role in New Zealand’s science-led response to the SARS-CoV-2 pandemic, identifying links between cases in community outbreaks to help stamp out local transmission⁵⁴. The Delta outbreak not only changed New Zealand’s strategy for controlling SARS-CoV-2, but it also changed the role of genomics. Much like other parts of the world, genomic surveillance of SARS-CoV-2 should continue to inform public health responses in New Zealand, but at a different granularity.

Supplementary Material.

Supplementary Table 1. List of GISAID accessions for genomes used in this analysis.

Supplementary Figure 1. Alignment of sequencing reads from isolate 21CV0515 to the Wuhan hu-1 reference genome. Sequencing reads are depicted as grey boxes with 3 prime ends of each read ending in an arrowhead. Deletions relative to the reference genome are shown as discontinuous regions in each read. The priming sites used in the sequencing protocol are shown in red. The fact sequencing reads containing the 10-nucleotide deletion span the deletion site, and the deletion site is not associated with a priming site are strong evidence that the deletion is genuine, and not an assembly artefact.

References.

- 1 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).
- 2 Dhar, M. S. *et al.* Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science* **374**, 995-999, doi:10.1126/science.abj9932 (2021).
- 3 Lopez Bernal, J. *et al.* Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N Engl J Med* **385**, 585-594, doi:10.1056/NEJMoa2108891 (2021).
- 4 Li, L. *et al.* Transmission and containment of the SARS-CoV-2 Delta variant of concern in Guangzhou, China: A population-based study. *PLOS Neglected Tropical Diseases* **16**, e0010048, doi:10.1371/journal.pntd.0010048 (2022).
- 5 Kenyon, G. Australia's struggle with the delta variant. *Lancet Infect Dis* **21**, 1358, doi:10.1016/s1473-3099(21)00579-x (2021).
- 6 Reading, P. C. & Strugnell, R. A. COVID-19 in Fiji. *Microbiology Australia* **42**, 192-195 (2021).
- 7 Lim, W. Y. *et al.* First nosocomial cluster of COVID-19 due to the Delta variant in a major acute care hospital in Singapore: investigations and outbreak response. *Journal of Hospital Infection* **122**, 27-34, doi:https://doi.org/10.1016/j.jhin.2021.12.011 (2022).
- 8 Chau, N. V. V. *et al.* An observational study of breakthrough SARS-CoV-2 Delta variant infections among vaccinated healthcare workers in Vietnam. *EClinicalMedicine* **41**, 101143, doi:10.1016/j.eclinm.2021.101143 (2021).
- 9 Shy, C.-G. *et al.* Rapid control of a SARS-CoV-2 B.1.617.2 (Delta) variant COVID-19 community outbreak: the successful experience in Pingtung County of Taiwan. *International Journal of Environmental Research and Public Health* **19**, 1421 (2022).
- 10 Jefferies, S. *et al.* COVID-19 in New Zealand and the impact of the national response: a descriptive epidemiological study. *The Lancet Public Health* **5**, e612-e623, doi:https://doi.org/10.1016/S2468-2667(20)30225-5 (2020).
- 11 Douglas, J. *et al.* Real-time genomics for tracking Severe acute respiratory syndrome coronavirus 2 border Incursions after virus elimination, New Zealand. *Emerg Infect Dis* **27**, 2361-2368, doi:10.3201/eid2709.211097 (2021).
- 12 Kung, S., Doppen, M., Black, M., Hills, T. & Kearns, N. Reduced mortality in New Zealand during the COVID-19 pandemic. *Lancet* **397**, 25, doi:10.1016/s0140-6736(20)32647-7 (2021).
- 13 Summers, J., Baker, M. G. & Wilson, N. *Mortality declines in Aotearoa NZ during the first two years of the Covid-19 pandemic*, <<https://blogs.otago.ac.nz/pubhealthexpert/mortality-declines-in-aotearoa-nz-during-the-first-two-years-of-the-covid-19-pandemic/>> (2022).
- 14 Baker, M. G. *et al.* The next phase in Aotearoa New Zealand's COVID-19 response: a tight suppression strategy may be the best option. *N Z Med J* **134**, 8-16 (2021).
- 15 Blair, A., de Pasquale, M., Gabeff, V., Rufi, M. & Flahault, A. The end of the elimination strategy: decisive factors towards sustainable management of COVID-19 in New Zealand. *Epidemiologia* **3**, 135-147 (2022).

- 16 Geoghegan, J. L. *et al.* Use of genomics to track coronavirus disease outbreaks, New Zealand. *Emerg Infect Dis* **27**, 1317-1322, doi:10.3201/eid2705.204579 (2021).
- 17 Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols* **5**, doi:10.1093/biomethods/bpaa014 (2020).
- 18 Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**, 33-46, doi:https://doi.org/10.1002/gch2.1018 (2017).
- 19 Aksamentov, I., Roemer, C., Hodcroft, E. B. & Neher, N. A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software* **6**, 3773 (2021).
- 20 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).
- 21 Hasegawa, M., Kishino, H. & Yano, T.-a. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174, doi:10.1007/BF02101694 (1985).
- 22 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587-589, doi:10.1038/nmeth.4285 (2017).
- 23 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518-522, doi:10.1093/molbev/msx281 (2017).
- 24 Zhou, Z. *et al.* GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res* **28**, 1395-1404, doi:10.1101/gr.232397.117 (2018).
- 25 Omasits, U., Ahrens, C. H., Müller, S. & Wollscheid, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**, 884-886, doi:10.1093/bioinformatics/btt607 (2014).
- 26 Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences* **110**, 228-233, doi:10.1073/pnas.1207965110 (2013).
- 27 Bouckaert, R. *et al.* An efficient coalescent epoch model for Bayesian phylogenetic inference. *Systematic Biology* in press (2022).
- 28 Douglas, J., Zhang, R. & Bouckaert, R. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLOS Computational Biology* **17**, e1008322, doi:10.1371/journal.pcbi.1008322 (2021).
- 29 Gowrisankar, A., Priyanka, T. M. C. & Banerjee, S. Omicron: a mysterious variant of concern. *Eur Phys J Plus* **137**, 100-100, doi:10.1140/epjp/s13360-021-02321-y (2022).
- 30 Redondo, N., Zaldívar-López, S., Garrido, J. J. & Montoya, M. SARS-CoV-2 Accessory proteins in viral pathogenesis: knowns and unknowns. *Frontiers in Immunology* **12**, doi:10.3389/fimmu.2021.708264 (2021).
- 31 Eichler, N. *et al.* Transmission of Severe acute respiratory syndrome coronavirus 2 during border quarantine and air travel, New Zealand (Aotearoa). *Emerg Infect Dis* **27**, 1274-1278, doi:10.3201/eid2705.210514 (2021).
- 32 Geoghegan, J. L. *et al.* Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *Nature Communications* **11**, 6351, doi:10.1038/s41467-020-20235-8 (2020).
- 33 Douglas, J. *et al.* Phylodynamics reveals the role of human travel and contact tracing in controlling COVID-19 in four island nations. *Virus Evolution*, **7**, 2, veab052 (2021).
- 34 Yoshimoto, F. K. The proteins of Severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *The Protein Journal* **39**, 198-216, doi:10.1007/s10930-020-09901-4 (2020).
- 35 Arya, R. *et al.* Structural insights into SARS-CoV-2 proteins. *Journal of Molecular Biology* **433**, 166725, doi:https://doi.org/10.1016/j.jmb.2020.11.024 (2021).

- 36 Cao, Z. *et al.* Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. *Cellular & Molecular Immunology* **18**, 746-748, doi:10.1038/s41423-020-00603-6 (2021).
- 37 Zhou, Z. *et al.* Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience* **24**, 102187, doi:<https://doi.org/10.1016/j.isci.2021.102187> (2021).
- 38 Lee, J.-G. *et al.* Characterization of SARS-CoV-2 proteins reveals Orf6 pathogenicity, subcellular localization, host interactions and attenuation by Selinexor. *Cell & Bioscience* **11**, 58, doi:10.1186/s13578-021-00568-7 (2021).
- 39 Finkel, Y. *et al.* The coding capacity of SARS-CoV-2. *Nature* **589**, 125-130, doi:10.1038/s41586-020-2739-1 (2021).
- 40 Tan, Y.-J. *et al.* Overexpression of 7a, a protein specifically encoded by the severe acute respiratory syndrome coronavirus, induces apoptosis via a caspase-dependent pathway. *J Virol* **78**, 14043-14047, doi:10.1128/JVI.78.24.14043-14047.2004 (2004).
- 41 Schaecher, S. R., Touchette, E., Schriewer, J., Buller, R. M. & Pekosz, A. Severe acute respiratory syndrome coronavirus gene 7 products contribute to virus-induced apoptosis. *J Virol* **81**, 11054-11068, doi:10.1128/JVI.01266-07 (2007).
- 42 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**, e00711-00720, doi:10.1128/JVI.00711-20 (2020).
- 43 Tse, H. *et al.* Genome Sequences of Three SARS-CoV-2 ORF7a deletion variants obtained from patients in Hong Kong. *Microbiol Resour Announc* **10**, doi:10.1128/mra.00251-21 (2021).
- 44 Foster, C. S. P. & Rawlinson, W. D. Rapid spread of a SARS-CoV-2 Delta variant with a frameshift deletion in ORF7a. *medRxiv*, 2021.2008.2018.21262089, doi:10.1101/2021.08.18.21262089 (2021).
- 45 Panzera, Y. *et al.* Consecutive deletions in a unique Uruguayan SARS-CoV-2 lineage evidence the genetic variability potential of accessory genes. *PLOS ONE* **17**, e0263563, doi:10.1371/journal.pone.0263563 (2022).
- 46 Pyke, A. T. *et al.* Replication kinetics of B.1.351 and B.1.1.7 SARS-CoV-2 variants of concern including assessment of a B.1.1.7 mutant carrying a defective ORF7a gene. *Viruses* **13**, 1087, doi:10.3390/v13061087 (2021).
- 47 Mazur-Panasiuk, N. *et al.* Expansion of a SARS-CoV-2 Delta variant with an 872 nt deletion encompassing ORF7a, ORF7b and ORF8, Poland, July to August 2021. *Euro Surveill* **26**, 2100902, doi:10.2807/1560-7917.ES.2021.26.39.2100902 (2021).
- 48 Joonlasak, K. *et al.* Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a. *Virus Res* **292**, 198233-198233, doi:10.1016/j.virusres.2020.198233 (2021).
- 49 Rosenthal, S. H. *et al.* Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. *bioRxiv*, 2020.2012.2010.418855, doi:10.1101/2020.12.10.418855 (2020).
- 50 Nemudryi, A. *et al.* SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Reports* **35**, 109197, doi:<https://doi.org/10.1016/j.celrep.2021.109197> (2021).
- 51 Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: Evolving the largest RNA virus genome. *Virus Res* **117**, 17-37, doi:<https://doi.org/10.1016/j.virusres.2006.01.017> (2006).
- 52 Stoye, J. P. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nature Reviews Microbiology* **10**, 395-406, doi:10.1038/nrmicro2783 (2012).
- 53 Douglas, J., Drummond, A. J. & Kingston, R. L. Evolutionary history of cotranscriptional editing in the paramyxoviral phosphoprotein gene. *Virus Evolution* **7**, doi:10.1093/ve/veab028 (2021).
- 54 Geoghegan, J. L., Moreland, N. J., Le Gros, G. & Ussher, J. E. New Zealand's science-led response to the SARS-CoV-2 pandemic. *Nature Immunology* **22**, 262-263, doi:10.1038/s41590-021-00872-x (2021).