

REGULAR ARTICLE

Analysis of SARS-CoV-2 haplotypes and genomic sequences during 2020 in Victoria, Australia, in the context of putative deficits in innate immune deaminase anti-viral responses

Robyn A. Lindley^{1,2,3}  | Edward J. Steele^{3,4} 

¹GMDxgen Pty Ltd, Melbourne, Victoria, Australia

²Department of Clinical Pathology, The Victorian Comprehensive Cancer Centre, Faculty of Medicine, Dentistry & Health Sciences, University of Melbourne, Melbourne, Victoria, Australia

³Melville Analytics Pty Ltd, Melbourne, Victoria, Australia

⁴CYO'Connor ERADE Village Foundation, 24 Genomics Rise, Piara Waters, Australia

Correspondence

Edward J. Steele, Melville Analytics Pty Ltd, Level 2, 517 Flinders Lane, Melbourne, VIC 3000 Australia.
Email: e.j.steele@bigpond.com

Funding information

Supported entirely by in-house company funding (GMDxgen Pty Ltd and Melville Analytics PTY Ltd)

Abstract

The SARS-CoV-2 epidemic infections in Australia during 2020 were small in number in epidemiological terms and are well described. The SARS-CoV-2 genomic sequence data of many infected patients have been largely curated in a number of publicly available databases, including the corresponding epidemiological data made available by the Victorian Department of Health and Human Services. We have critically analysed the available SARS-CoV-2 haplotypes and genomic sequences in the context of putative deficits in innate immune APOBEC and ADAR deaminase anti-viral responses. It is now known that immune impaired elderly co-morbid patients display clear deficits in interferon type 1 (α/β) and III (λ) stimulated innate immune gene cascades, of which APOBEC and ADAR induced expression are part. These deficiencies may help explain some of the clear genetic patterns in SARS-CoV-2 genomes isolated in Victoria, Australia, during the 2nd Wave (June–September, 2020). We tested the hypothesis that predicted lowered innate immune APOBEC and ADAR anti-viral deaminase responses in a significant proportion of elderly patients would be consistent with/reflected in a low level of observed mutagenesis in many isolated SARS-CoV-2 genomes. Our findings are consistent with this expectation. The analysis also supports the conclusions of the Victorian government's Department of Health that essentially one variant or haplotype infected Victorian aged care facilities where the great majority (79%) of all 820 SARS-CoV-2 associated deaths occurred. The implications of our data analysis for other localized epidemics and efficient coronavirus vaccine design and delivery are discussed.

1 | INTRODUCTION

The coronavirus pandemic due to SARS-CoV-2 has engulfed the world and has exacted enormous health and economic tolls. Since it first emerged in Wuhan, China, in late

November and early December 2019,¹⁻³ the health impact has been mainly on our elderly co-morbid citizens. In designing an effective vaccine, it is important, in our view, to understand how this pandemic coronavirus genetically diversifies at the RNA and protein levels as it infects different

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Scandinavian Journal of Immunology* published by John Wiley & Sons Ltd on behalf of The Scandinavian Foundation for Immunology.

human populations. This is particularly the case for the elderly co-morbid subgroup where the deficits in innate immunity are clearly evident⁴⁻⁸ as well as compromised adaptive immunity.^{9,10} Type I and type III interferon (IFN) inducible anti-viral immunity is particularly effected.¹¹ Longitudinal studies show that patients in this subgroup appear 'immune defenceless' to coronavirus respiratory tract infections and are at a very high risk for severe outcomes including death.¹¹ Important reviews on the special susceptibility of this aged co-morbid group have appeared.^{10,12} An integrated schematic of the innate immune deficits in relation to adaptive immune responses and severity of viral replication in healthy and susceptible human subjects has been published by Sette and Crotty¹⁰ which clearly shows the barrier of induced innate immunity which normally quells viral replication in infected healthy subjects in the first few days of infection. Such a picture is completely consistent with a very large population-level observational study in Denmark (4 million PCR-tested individuals) showing that the elderly are more often immunologically vulnerable to reinfection with SARS-CoV-2.¹³

We have started the process of investigating exactly how SARS-CoV-2 varies genetically to gain insight into more effective vaccine design. In a previous study, we analysed the full-length genomic sequences generated over the first three months of the pandemic from late December 2019 through January 2020 in China (mainly Wuhan), then via the hot spot zones in Spain, some early sporadic outbreaks in California and Washington State (through February 2020) and then to the large-scale explosive outbreak in New York City in mid to late March 2020.¹⁴ In that study, we used a haplotype rather than the more mainstream phylogenetic tree approach to the temporally emerging SARS-CoV-2 sequence data¹⁵⁻¹⁷ The haplotype approach directly compares the full-length test sequence aligned against the Hu-1 reference. This type of comparison provides insights into the putative APOBEC and ADAR deaminase-driven cytosine to uracil (C > U) and adenosine-to-inosine (A > I/G) variation mechanisms per se which are not necessarily obvious or available with the phylogenetic approach of the Pangolin classification system¹⁵⁻¹⁷ (which generates putative person-to-person or P-to-P lineages, clusters and clades). Indeed, RNA-sequence haplotypes, in our view, are functional strings implying stable RNA secondary structures required for maximal replicative efficiency in that host's genetic/biochemical background.¹⁴ The other implication is that new vaccine designs and therapies may need to specifically target functionally important secondary structures with the presumptive replicative stability of the viral RNA genome (eg reviewed in Robson et al¹⁸).

Our detailed haplotype and sequence analyses reported here further extend the main findings from the previous

work¹⁴ but now within a small well defined geographical urban setting. The main findings show how a single viral clone or haplotype of SARS-CoV-2, either unmutated or lightly mutated, can dominate a local epidemic and cause considerable health impacts on elderly citizens who are putatively immune defenceless. The person-to-person (P-to-P) spread of the same haplotype clone in closed institutions and between institutions was most likely affected by more healthy asymptomatic carriers of the disease. We are thus testing the hypothesis that due to *expected deficits* in type I and type III IFN inducible innate anti-viral immunity¹¹ involving activated APOBEC and ADAR deaminase expression¹⁹⁻²¹ there will be a lowered level of observed mutagenesis in isolated SARS-CoV-2 genomes isolated from such patients. As we discussed earlier¹⁴ the anti-viral properties of the induced innate immune APOBEC and ADAR deaminases are well documented in the literature. They are powerful viral genome mutators causing predominantly transition mutations such as C > T / G > A and A > G/T > C single nucleotide variations (SNV) in single-stranded and double-stranded viral RNA and DNA genomes.²¹⁻²⁵ Such mutations at some deaminase sites can be pro-viral.²¹ On occasion, such mutators can also go off target affecting host genomes and thus contributing to cancer mutagenesis as recently reviewed by Lindley.²⁶

Here, we report on an exhaustive analysis of the levels of mutagenesis in full-length SARS-CoV-2 genomes that have been, due to the mass testing coverage in Victoria, isolated mainly from putative elderly infected patients (and their carers and tending health professionals according to media and government reports) in order to assess a fit with the hypothesis that a predicted lowered innate immune APOBEC and ADAR anti-viral deaminase responses in a significant proportion of elderly patients would be consistent with/reflected in a low level of observed mutagenesis in many isolated SARS-CoV-2 genomes. The data, in an epidemiologically somewhat unique cohort (isolated island community, excellent biomedical research environment, first class public health response system), are consistent with our stated hypothesis.

2 | MATERIALS AND METHODS

2.1 | Publicly available data resources

Supporting Information Files A, B, C, D, E, F, G summarize much of this publicly available data and information and are cited below.

- a. *Victorian Government Website* : Summary statistics of COVID-19 new cases and whether they were acquired in Australia (ie in Melbourne or Victorian regions)

from a known source or from an unknown source have been recorded by the Government of Victoria at the Department of Health website <https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data> and in the *Herald Sun* newspaper, Figures 1, 2, 3, 4 and 5 and Table 1, and Supporting Information File A and F summarizes these infection data. Clearly as Table 1 shows most elderly people who died in Victoria (97.8%) on contracting SARS-CoV-2 were over 60 years, of median age 80–89 years, and these were concentrated in aged care facilities (Supporting Information File F). However, not all infected resident cases in aged care died; yet it was a significant yet variable fraction across a range of institutions, approximately 30%–40% of resident cases (Supporting Information File F).

b. *Herald Sun newspaper*: A full page of SARS-CoV-2 infection statistical data and detailed Victorian State map of case incident summaries (from mid-July to October 28) appeared each day in this Melbourne daily tabloid newspaper. Many of these pages were preserved and a scanned (pdf) digital record created. Examples for late July and early August—just before hard Stage 4 lock downs began on August 2—are shown in Supporting Information File A. Data in this newspaper record were used to construct the summary in Figure 2 from data in Supporting Information File C. This page of the newspaper also had a story of the day discussing the main challenges which the health authorities confronted. In our considered view, the most important reports, and there were many, were those on the very high incidence of ‘Community Transmissions’. These cases, which are defined by rigorous epidemiological contact tracing and genomic sequencing linked to a known cluster, are defined as SARS-CoV-2 infections

acquired from an unknown source (Figures 2, 3, 4 and 5, and Supporting Information file A, and see ‘COVID-19 Hotel Quarantine Inquiry, Final Report and Recommendations’²⁷). These cases were largely absent or rare in the Victorian 1st Wave where all secondary cases acquired in Australia came largely from infected overseas travellers into Melbourne. The numerous ‘Community Transmission’ cases in the 2nd Wave largely drove the Victorian Government’s social distancing lock down policies. During the period from August 1 to September 6, these numbers went from highs of 94–174 per day (recorded across the City and the State) to 55–8 per day by September 6. (Supporting Information File C). They eventually declined as a significant driving source of infections as the epidemic petered out. The numbers per day are listed here to indicate they are small enough in daily aggregates for the teams of contract tracing units across the state to manage. However, the phrase ‘Community Transmissions’ quickly evolved over many detailed news reports of joint press conferences (Supporting Information File A) by the Chief Victorian Health Officer (Professor Brett Sutton) and the Premier of Victoria (Hon. Daniel Andrews) to become genuine ‘Mystery Cases’—as they could not be traced to an acquisition from a known contact case, that is no patient zero, or linked to a known genomic sequence cluster.

a. *GenBank/NCBI Virus*. This publicly accessible site is hosted at The National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH), Bethesda, Maryland, USA, which curates SARS-CoV-2 sequences at <https://www.ncbi.nlm.nih.gov/sars-cov-2/#nucleotide-sequences>. This was also used

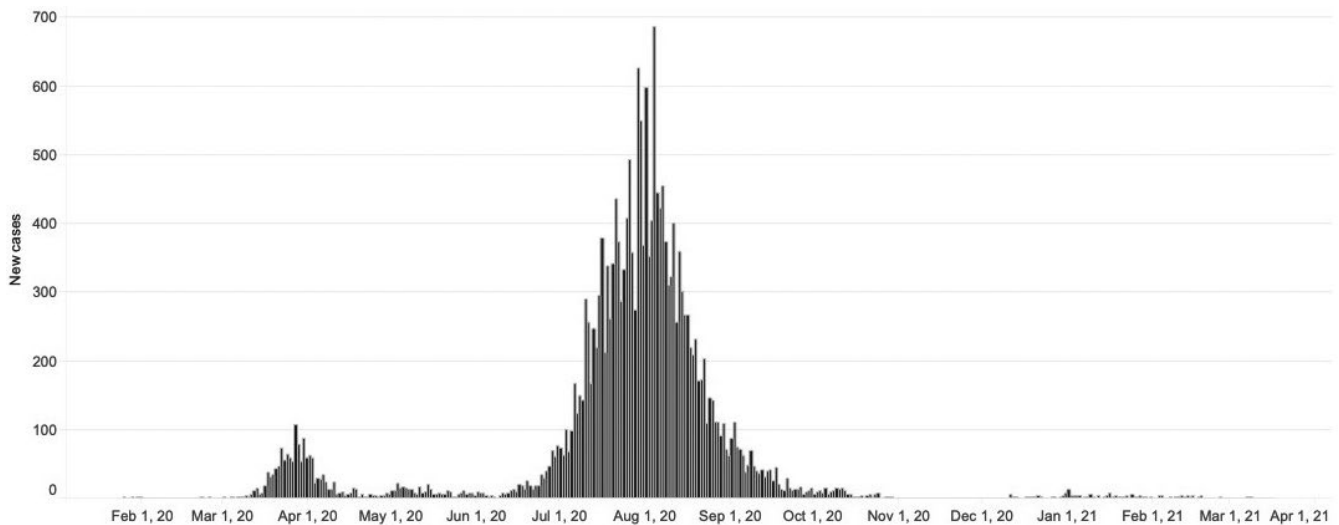


FIGURE 1 New SARS-CoV-2 cases per day recorded in Victoria, Australia, during 2020. These data can be accessed at <https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data>

for all sequences haplotyped and analysed in Steele and Lindley.¹⁴ As before, the simple and straight forward NCBI Virus alignment tool at the site was used for all sample alignments against the original Wuhan Hu-1 reference sequence (NC_045512.2). The data for the haplotyping screens of 96 alignments of variable size can be found in the Supporting Information File B. It has sheets showing

TABLE 1 People who have passed away with coronavirus (COVID-19) in Victoria throughout 2020

Age Group	Male	Female	Total
00-09	0	0	0
10-19	0	0	0
20-29	1	0	1
30-39	2	0	2
40-49	1	0	1
50-59	10	4	14
60-69	20	9	29
70-79	83	46	129
80-89	166	182	348
90+	107	189	296
Total	390	430	820

Notes: Data from Victorian Department Health and Humans Services website as 13 March 2021. <https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data>

Numbers of deaths of all SARS-CoV-2 cases in Victoria 2020. This is from the Victorian Government Department of Health website as 17 March 2021. The median age at death assumed by COVID-19 is 80-89 years, and 97.8% of all deaths were in patients >60 years; and 94.3% are aged >70 years.

<https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data>

the alignment number and sequence IDs; the alignments grouped by collection period from patients; a summary of numbers of various haplotypes versus collection dates; tables showing accumulated haplotype numbers over different collection times. An excel table and histogram plots (see Figures 6 and 7) showing the emergence of the almost clonal L241f.1vic haplotype over time versus the related L241f.1 haplotype in collections from early March 2020 to early September 2020 is assembled from data in Supporting Information File B and D and also presented in Supporting Information File E.

An important caveat is there is no metadata available in the public domain linking a particular patient's clinical/epidemiological data to a given GenBank Accession number for a full-length SARS-CoV-2 genomic sequence. All associations are thus inferred from the sheer dominance of all main SARS-CoV-2 case data (and deaths) amongst the elderly and within aged or nursing care facilities (Supporting Information F). Inquiries in writing were made to The Victorian Department of Health and The Peter Doherty Institute, and these were unsuccessful in securing metadata information.

2.2 | Analysis of SARS-CoV-2 sequences at NCBI virus

a. Main Haplotype Screen. The haplotype analyses using the NCBI Virus 'Align' tool screened 12,798 SARS-CoV-2 genomes collected in Victoria from first travellers into Melbourne from Wuhan and other world cities from January through to October 2 (Supporting Information

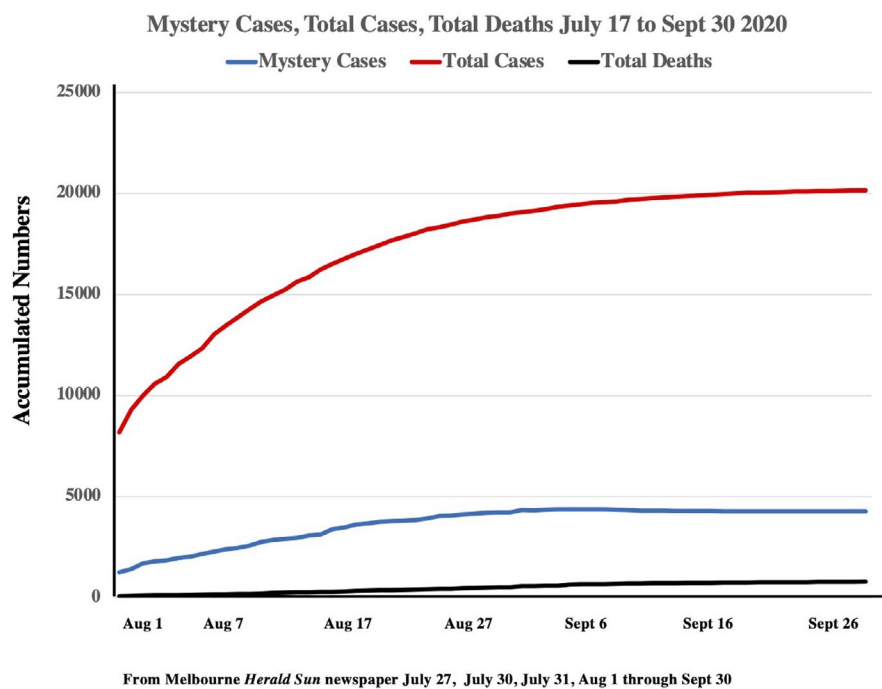
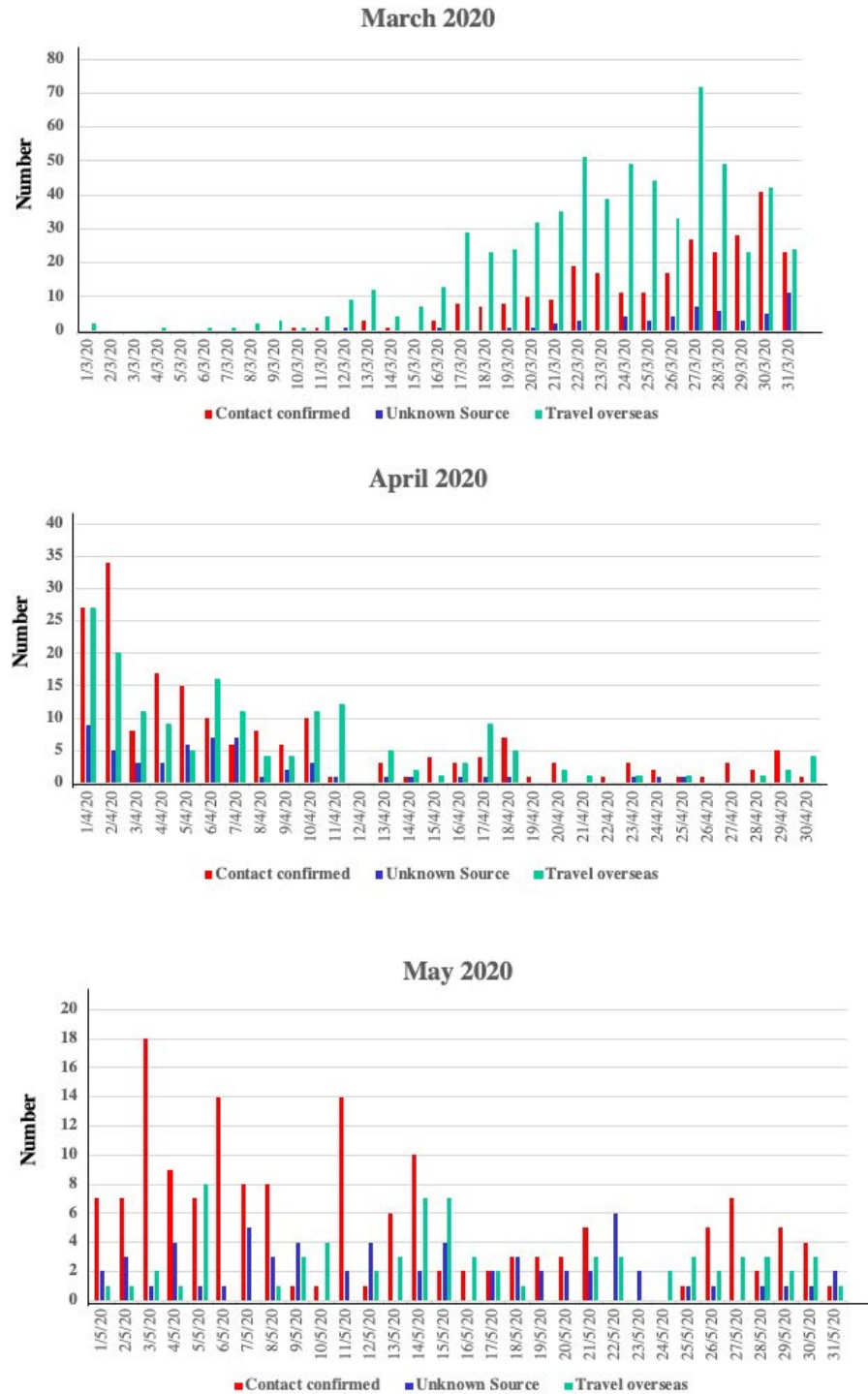


FIGURE 2 Total accumulated mystery cases, total cases and total accumulated deaths from July 27 through September 30—as published daily in the *Herald Sun* newspaper (Supporting Information File C)

FIGURE 3 Numbers of confirmed cases acquired versus cases acquired from unknown cause sin Victoria 1st wave March through May 2020. Primary data in NCOV_COVID_Cases_by_Source_20210131.csv at: <https://www.dhhs.vic.gov.au/victorian-coronaviruss-covid-19-data>



File B). The haplotype markers across the Hu-1 SARS-CoV-2 genome are those shown in Table 2 which includes the L241f.1vic haplotype which was defined more clearly in the detailed later Sequence Alignment analyses of all L241f.1 haplotypes from March, April, May and June and sample collection times in July, August and September (Supporting Information File D). Table 2 is thus a refinement and revised update of the original provisional haplotyping table.¹⁴ The main addition is the L241f.1 haplotype with the distinctive SNV tandem

changes GGG>AAC at p.28881-3. This was in international circulation as it was observed at low frequency in New York City (late March 2020)¹⁴ and also in many collections in Florida through April, May, June and July 2020 (data not shown).

Summaries of all the data on the haplotype screen are in the Supporting Information File B, Excel Book Sheets S1-S5. This large haplotype screen employed SNV markers at p.241, p.1059, p.3037, p.8782, p.11083, 14 408, p.25563, p.28144, p.28881-3.

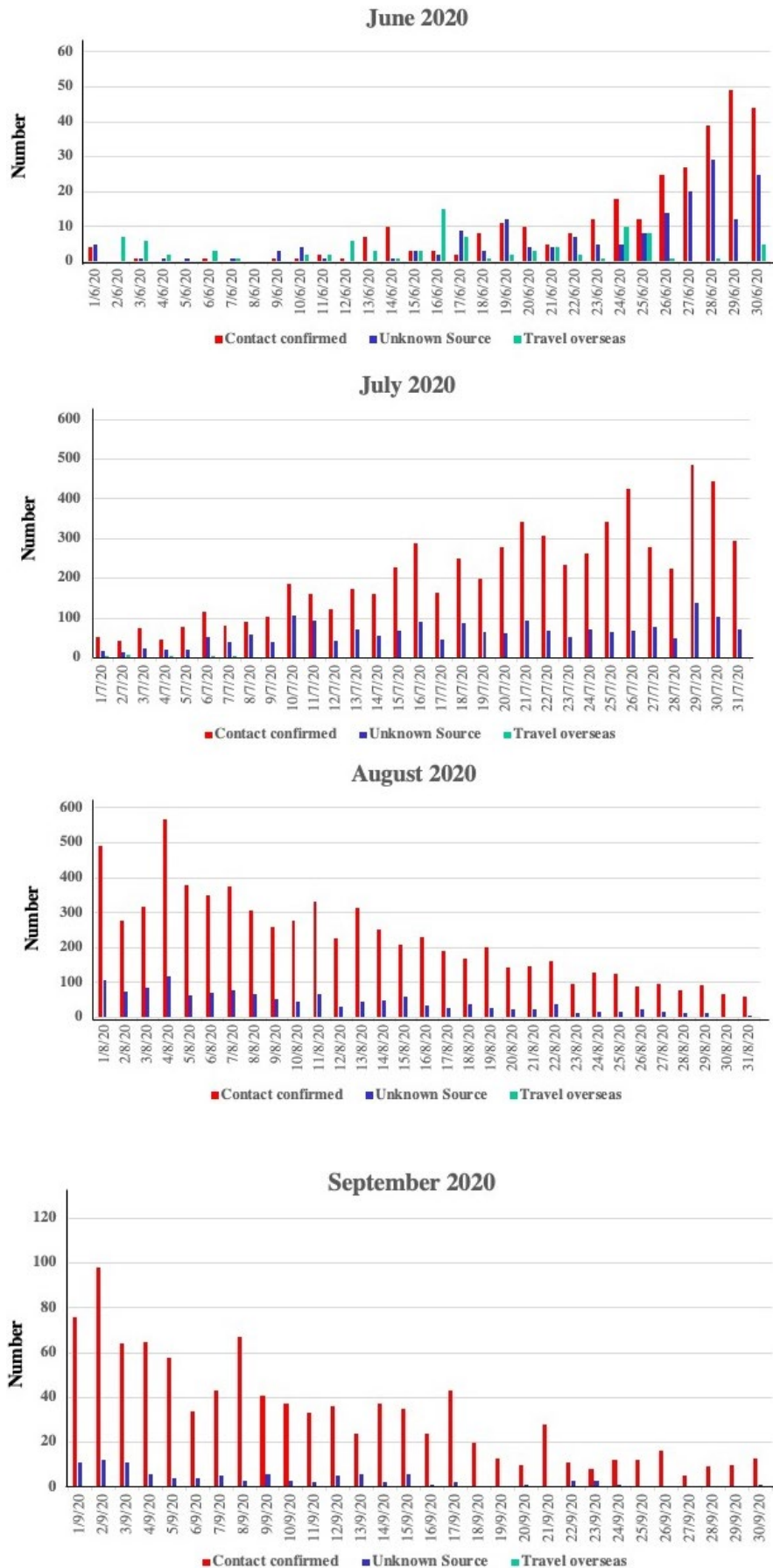
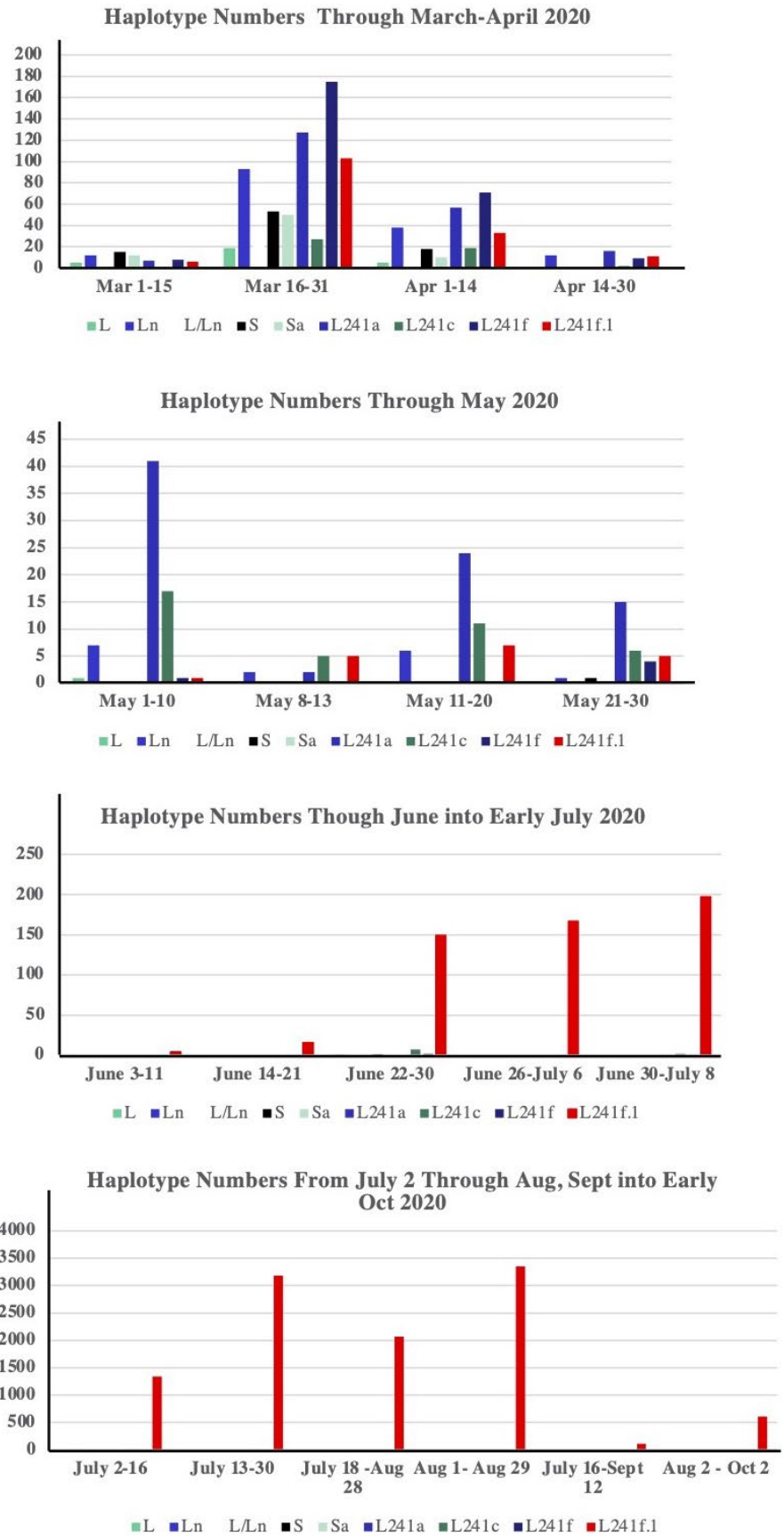


FIGURE 4 Numbers of confirmed cases acquired versus cases acquired from unknown cause in Victoria 2nd wave June through August 2020. Primary data in NCOV_COVID_Cases_by_Source_20210131.csv at: <https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data>

FIGURE 5 Numbers of confirmed cases acquired versus cases acquired from unknown cause in Victoria 2nd wave tail September 2020. Primary data in NCOV_COVID_Cases_by_Source_20210131.csv at: <https://www.dhhs.vic.gov.au/victorian-coronavirus-covid-19-data>

FIGURE 6 Haplotype numbers by collection period in first screen from data at NCBI virus. The primary summarized data are in Supporting Information File B, and these plots are from Sheet S4



b. Genomic Sequence Alignments of L241f.1 haplotypes March, April, May, June and for Selected Collection Times July, August and September. Full-length genomic

sequence analyses of L241f.1 haplotypes with collection times were conducted (Supporting Information File D) where the full patterns of SNV variability across many

Numbers L241f.1 and L241f.1vic Sequences Sampled March through September 2020 at NCBI Virus

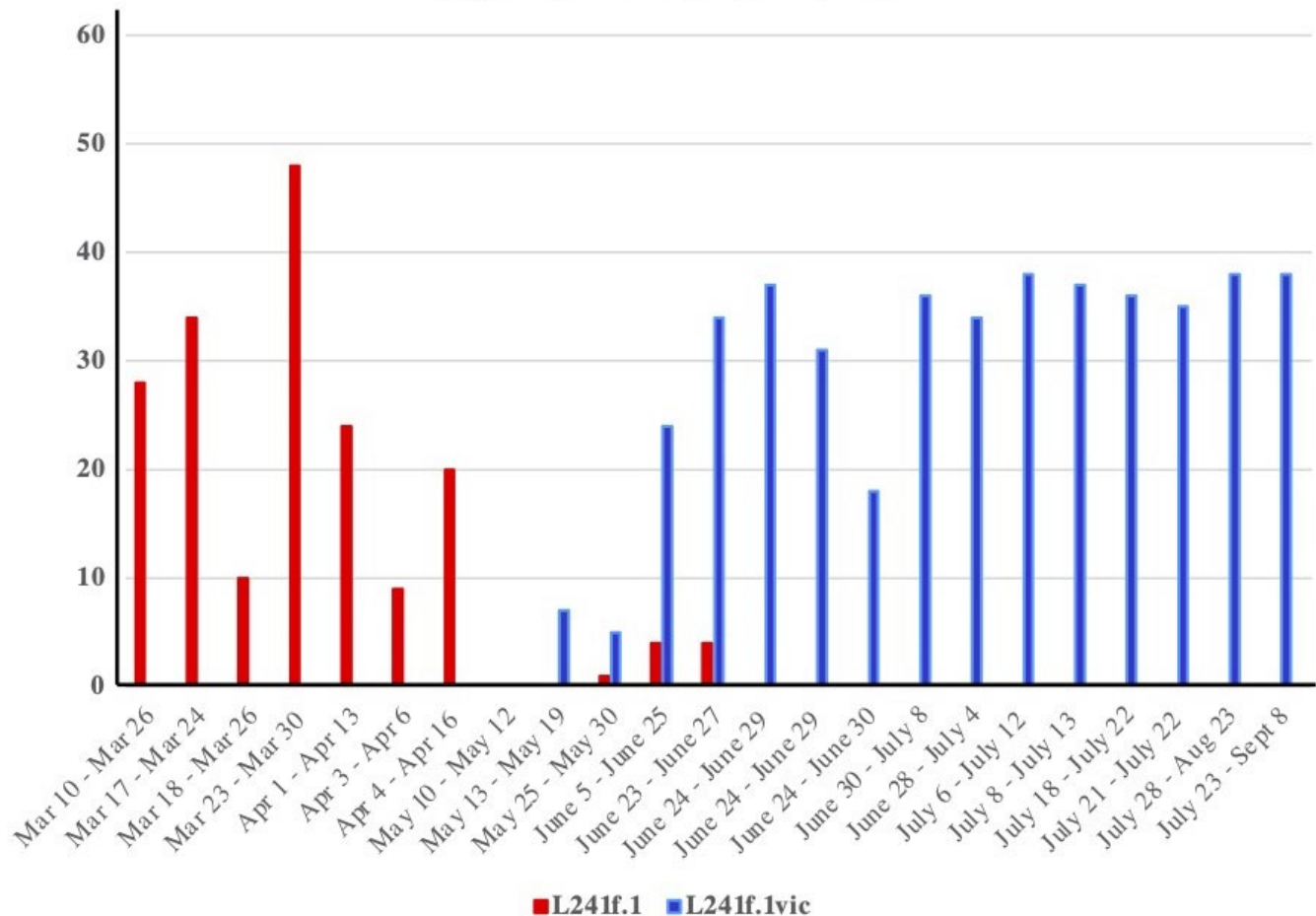


FIGURE 7 Emergence of L241f.1vic haplotypic sequence and disappearance of the L241f.1 haplotypic sequence from the uploads at GenBank/NCBI Virus. From Supporting Information Files D and E

collection time points can be inspected in that wider data set. The alignments discussed here were preserved as screen shot records of each NCBI Virus alignment for later construction of Variable Site Diagrams, or VSDs which clarify haplotype analysis as discussed¹⁴ (Supporting Information File D). The putative impact of SNV changes on protein structure and function was qualitatively assessed as in Steele and Lindley¹⁴ using the colour code for amino acid conservation or change in Supporting Information File D.

The VSDs were sorted by sub-variants of the L241f.1 haplotype and also by sequences that were clearly unmutated, usually grouped at the bottom of each VSD figure. In this way, lightly mutated haplotypes can also be easily identified and P-to-P transfers and P-to-P groups can also be readily identified. The SNV patterns from VSDs for the L241f.1vic haplotype are summarized in Table 3.

2.3 | Independent haplotype analysis of 'Vic' tagged SARS-CoV-2 sequences uploaded to GISAID database

An independent check was kindly conducted by Dr Jared Mamrot on the above haplotype screen and sequence alignment analyses on Victoria, Australia SARS-CoV-2 sequence data at NCBI Virus. The catalogue of Victorian SARS-CoV-2 sequences at the GISAID db (with GenBank ID links) was downloaded and screened for haplotype numbers using the haplotype markers defined in Table 2 at given collection dates spanning the 2nd wave, June 1 to September 30. Specifically, all SARS-CoV-2 viral genomes were obtained from GISAID (<https://www.gisaid.org/>) on 15 March 2021 and from GenBank (<https://www.ncbi.nlm.nih.gov/sars-cov-2/>) on 23 March 2021. FASTA sequences with 'Vic' in the identifier (case-insensitive search) were extracted from both databases, combined, and duplicates were removed.

Sequences collected between 01/06/2020 and 30/09/2020 were extracted ($n = 12,009$) and aligned to the NC_045512.2 SARS-CoV-2 reference genome using MAFFT (v7.475; <https://mafft.cbrc.jp/alignment/software/>) with options ‘--6merpair --keeplength --addfragments’. Haplotypes were defined according to 19 genomic positions, relative to the reference genome as listed in Table 2: p.241, p.1059, p.3037, p.1163, p.7540, p.8782, p.11083, p.14408, p.16647, p.17747, p.17858, p.18060, p.18555, p.22992, p.23401, p.23403, p.25563, p.28144, p.28881-3. Nucleotides at these positions were identified for each sample, collated and summarized. These data are in Supporting Information File G and for further ease of principle haplotype identification and analysis were sorted first by column T (p.28881-3), then by column F (p.7540), then by column C (p.1059) and then by column D (p.3037).

2.4 | Correspondence between Pangolin classification and main haplotypes in Table 2

At time of writing, we are in the process of forming a translation table for the main Steele-Lindley (S/L) ‘replicative’ haplotypes (Table 2) with the PANGO classification system’. At this stage it is evident that.

PANGO ‘B’ = L Hap of S/L = Wuhan ref seq December 2019 (Hu-1).

PANGO ‘P.1 and B.1.1.7 and B.1.1.348’ = L241f.1 haplotype variants of S/L. This would cover the ‘UK Mutant’ and the different ‘Vic Mutant’ (L241f.1vic) that infected many Melbourne aged care and nursing home and presumed secondary asymptomatic contacts.

PANGO ‘B.1 and B.1.526 and B.1.429 and B.1.526.1’ = L241a of S/L. A common and frequent haplotype seen in overseas air travellers in the Vic 1st Wave March-April 2020.

PANGO ‘B.1.525’ = L241f of S/L. A common and frequent haplotype seen in overseas air travellers in the Vic 1st Wave March–April 2020.

Therefore, a number of PANGO lineages in this first small sample tested can be understood in terms of the sequence architecture of the main coordinated positional changes in the 29 903 nt Hu-1 RNA sequence (which is ‘B’ or ‘L’). Evidently, PANGO is tracking human passaged variants in local global regions and they appear as Sub-Haplotype variants of the main Haplotypes we have recorded in the first three months of the pandemic.¹⁴ In other NCBI Virus, multiple alignment analyses of pairs or groups of the same PANGO variants collected at different times in different geographical regions it is apparent to us these PANGO variants are the end products of multiple P-to-P transfers accumulating largely deaminase-mediated¹⁴ mutations at each transfer. Estimates of the number of putative

mutagenic P-to-P transfers away from the primary haplotypes reported in Steele and Lindley¹⁴ range from 4 P-to-P transfers for B.1.1.7 to 6 or 7 transfers to generate B.1.525 and 3 or 4 transfers to generate B.1.526 lineages (data not shown, Steele EJ, Gorczynski RM et al in preparation).

3 | RESULTS

The opportunity to gain a deeper insight into the SARS-CoV-2 genetic variation mechanisms and viral adaptation strategies during the pandemic was offered by the definitive sets of observations on SARS-CoV-2 case numbers in Victoria, (mainly in Melbourne) Australia. The Victorian epidemics are scientifically definitive for several reasons:

1. The genesis of the two epidemics were clearly different, the 1st Wave due solely to overseas travellers entering Australia from Northern Hemisphere infected zones, the 2nd Wave to infections clearly acquired in Australia (Figures 1, 3, 4, 5). This is a very important demarcation in the case data.
2. The health authorities implemented strict lock downs and social distancing protocols, including mandatory mask wearing and mass PCR testing as well as quarantining of overseas travellers on arrival.
3. The health authorities identified hot zones all over the state and convinced tens of thousands of citizens to get PCR-tested on a mass scale whether showing symptoms or not (upwards of 30,000 PCR tests on oro-nasal swabs per day through May, June, July, August and September 2020).
4. Teams of contact tracers were actively coordinated throughout Melbourne and regions and were largely comprehensive despite some failures and much public political criticism of the regimentation of public behaviour involved.

As a consequence, a large amount of epidemiological data, PCR testing data and full-length genomic sequences were generated and largely placed in the public domain. This allows the comprehensive analysis and conclusions we record here. These observations and genomic sequence data are represented in 2020 by the main 1st and 2nd Wave SARS-CoV-2 epidemics: the 1st Wave in March-April (with a May blip in new cases and secondary transmissions) and the 2nd Wave beginning mid-June through July, August and September (Figure 1). Thus, many of the full-length genomic sequences for these epidemics ($\geq 12,000$) are publicly available for direct analysis using the alignment tools at the NCBI GenBank/NCBI Virus website. Moreover, whilst these two epidemics in Australia have been very controversial and caused much political dysfunction in

TABLE 2 Haplotypes and main sites defining SARS-CoV-2 common strain variants updated and revised for Victoria -Corrected 30.6.21

AA class->	P<>NonP	SYN	P<>NonP	SYN	SYN	NonP<>NonP	NonP<>NonP	SYN	
	5'UTR	Thr<>Ile	Phe<>Phe	Ile<>Phe	Thr<>Thr	Ser<>Ser	Leu<>Phe	Pro<>Leu	Thr<>Thr
HAP	p.241	p.1059	p.3037	p.1163	p.7540	p.8782	p.11080/83	p.14408	p.16647
L (Hu-1)	C	C	C	A	T	C	G	C	G
Ln	C	C	C	A	T	C	T	C	G
L-241a	T	T	T	A	T	C	G	T	G
L-241b	T	T	C	A	T	C	G	T	G
L-241c	T	C	T	A	T	C	G	T	G
L-241d/s	T	T	T	A	T	C	G	C	G
L-241e	T	C	C	A	T	C	G	T	G
L-241f	T	C	T	A	T	C	G	T	G
L-241f.1	T	C	T	A	T	C	G	T	G
L-241f.1vic	T	C	T	<i>T</i>	<i>C</i>	<i>C</i>	G	T	<i>T</i>
L-241g	T	C	C	A	T	C	G	T	G
S	C	C	C	A	T	T	G	C	G
Sa	C	C	C	A	T	T	G	C	G
Sb	C	C	C	A	T	T	G	C	G
Ss	C	C	C	A	T	T	G	C	G

Notes: p.11080/83 is a pan-Haplotype marker (G/T) of putative 8oxoG non-deaminase modification at a W_G site that can occur across haplotypes.

Key changes in *italic* to focus on for Victorian 2nd wave analyses.

The table shows haplotypes and main sites defining SARS-CoV-2 common strain variants documented in China, USA, Spain (January-March) Florida (May, June, July), California (March), with Update and Revision for Victoria as 16.3.21 Post Publication Steele and Lindley.¹⁴

p.7540 site annotation corrected as 30.6.21 from P<>P Ser<>Thr tp SYN Thr<>Thr.

Victoria,²⁷ the epidemics nevertheless have features that allow succinct scientific analysis not necessarily available to Northern Hemisphere infected zones (where the great bulk of global infections have occurred and continue to occur at the time of writing).

The Results section thus highlights the statistics of the SARS-CoV-2 cases through 2020, the complete haplotyping of all publicly available SARS-CoV-2 genomes, and a detailed comparative full-length sequence analysis of all L241f.1genomes collected in March, April, May, June and sample sets selected at different times through early, mid and late July, and though August and September (all these alignments are summarized as VSD in Supporting Information File D, with SNV summary in Table 3).

3.1 | Overview SARS-CoV-2 case statistics of the 1st and 2nd waves in Victoria in 2020

The pandemic in Australia began with travellers into Australia from overseas infected zones, beginning in late

January 2020 and continuing through to the end of April and also later times into May 2020.²⁸ The numbers of New Cases per Day are shown in Figure 1.

The main 1st and 2nd Waves are shown—March–April, then from June through September and early October. There is a slight blip in May which can be considered for operational purposes the tail of the 1st Wave. The 1st wave was generated solely by infected travellers into Melbourne, and the 2nd Wave from mid-June through to the end of October generated almost exclusively by new cases acquired in Australia. About 79% of all cases acquired in Australia (ie Victoria) in the 2nd wave were by contact with an identifiable SARS-CoV-2 ‘case’ or by assignment to a known SARS-CoV-2 genomic sequence cluster. Many however were acquired from an unknown source and could not be assigned by contact tracing or linked to a known cluster by genomic sequencing (Figures 2-5).

Further insight into these case categories is provided at the Victorian Department of Health website in the data displayed in Figures 3-5 for March, April, May, June, July, August and September. This is a very informative set of plots based on recorded daily Acquired Cases in

TABLE 3 SNV frequency and P-to-P spread among L241f.1vic haplotype sequences 5 June through 8 September 2020

Collection Date 2020	No. of L241f.1vic Sequences	No. L241f.1vic Sequences with SNV Differences					No. P-to-P Clusters with Shared SNV Differences			
		0	1	2	3	≥4	1	2	3	≥4
June 7-June 25 (L241f.1 Align#1)	25	4	6	1	2	0	4	0	0	0
June 23-June 27 (L241f.1 Align#2)	34	9	1	5	0	0	3	0	0	0
June 24-June 29 (L241f.1 Align#3)	37	14	13	4	2	0	3	2	9	0
June 24-June 29 (L241f.1 Align#4)	31	7	3	6	0	0	3	0	0	0
June 24-June 30 (L241f.1 Align#5)	18	6	3	2	1	0	3	0	0	0
June 30-July 8 (L241f.1 Align#16)	36	4	1	8	3	0	4	0	0	0
June 28-July 4 (L241f.1 Align#13)	34	7	10	10	0	1	6	0	0	0
July 6-July 12 (L241f.1 Align#17)	38	13	13	5	4	3	6	0	0	0
July 8-July 13 (L241f.1 Align#18)	37	8	16	9	1	2	5	2	0	2
July 18-July 22 (L241f.1 Align#19)	36	12	12	7	4	1	5	0	2	0
July 21-July 22 (L241f.1 Align#20)	35	4	7	10	7	8	2	3	1	1
July 28-Aug 23 (L241f.1 Align#15)	38	4	9	10	10	5	6	2	0	0
July 23-Sept 8 (L241f.1 Align#14)	38	4	10	9	10	5	4	2	2	0

Note: Sequence collections times, alignment number from variable site diagrams in Supporting Information File D.

3.2 | Haplotype analyses of NCBI virus publicly available genomic sequencing data

The positional SNV markers for the first haplotype screen of 12,798 genomic sequences at NCBI Virus (versus the Hu-1 reference) were (Table 2) at p.241, p.1059, p.3037, p.8782, p.11083, p.14408, p.23403, p.25563, p.28144 and p.28881-3. Variants of the S haplotype (p.8782/p.28144) were further subtyped by markers at p.17747, p.18858 and p.18060. The marker p.11083 G/T we have noted as a pan-haplotype SNV but does distinguish L (Wuhan) from Ln (New York). In some cases in the aligned sequence data (Supporting Information File B), an N run across the

p.11083 site meant the haplotype was either L or Ln. L, Ln or L/Ln were common in the 1st Wave of travellers to Melbourne. A summary of all the haplotypes available at GenBank/NCBI Virus March 1 through October 2 is shown in Figure 6.

It is clear that there is considerable haplotype diversity through March and April. These haplotype patterns are consistent with the lineage and cluster patterns reported in Seemann et al²⁸ and in the sample we reported in Steele and Lindley.¹⁴ For our immediate purposes here, note that the L241f.1 haplotype was prominent, but not dominant, in March 2020 and was at low frequency in late April and into first week of May. These markers, defining the

L241f.1 sequence, then picked up sequence signals at low frequency from May 8 to 13 and maintained at low numbers right through to mid-June. Notice that by mid-June, most of the other haplotypes typical of the travellers entering Australia had dropped to low or undetectable numbers as numbers of international flights arriving in Australia were reduced.

Then from the third week of June, right through to the end of the collection period, (Oct 8) the L241f.1 haplotype (as defined Table 2) dominated all uploaded sequences at the NCBI Virus website.

It later became apparent when the actual full-length sequences of these L241f.1 haplotypes were examined in greater detail, that a clear new variant of L241f.1 had emerged from about May 10 with additional six coordinated changes (at largely riboswitch sites as defined earlier) at p.1163, p.7540, p.16647, p.18555, p.22992 and p.23401 now defining the L241f.1vic haplotype (as named by us here). This haplotype was the L241f.1vic variant that dominated all sequences uploaded to GenBank/NCBI Virus through to the end of the 2nd Wave (Figure 7). We have no clear sequence-linked metadata, but a plausible assumption is these were causally associated with the daily reported flareups of case numbers (see spikes in Figure 1, particularly through July) in the many Victorian SARS-CoV-2 case hotspots at aged care facilities across the inner and outer Western (and some Eastern) suburbs of Melbourne (Supporting Information File F). The flare ups in late June-early July are consistent with the well-publicized outbreaks in Melbourne in the nine tower blocks in the north west of the city in Kensington. All these high population density towers, with many three generation families from migrant communities, were ‘.. put into “hard” lockdown on 4 July, with 3000 residents told not to leave their homes for five days. State police was brought in to ensure compliance, and as community transmission continued to grow, lockdown measures were applied across Melbourne.’²⁹

3.3 | Full-length genomic sequence alignments of L241f.1 haplotypes of NCBI virus uploaded sequence data

We then set out to establish whether or not the set of SARS-CoV-2 full-length sequences of the L241f.1 haplotype, as defined by the markers in Table 2, is the same basic genomic sequence throughout the 1st and 2nd Waves. All the L241f.1 haplotypes available from March, April, May, June and at later sample collection times in July, August, September (we sampled groups across representative collection times as the numbers there were very large) were aligned against Hu-1 and the sequence

alignments presented as VSD as shown in Supporting Information File D, and SNVs analysed as summarized in Table 3 for these collection times. The full set of sample times is in Supplementary Information File D which shows a more detailed trend in the VSD patterns over time, particularly in the L241f.1vic haplotype that went on to dominate sequence uploads exclusively from mid-June. The type of SNV change at a given VSD position relative to the Hu-1 reference sequence (C > T, A > G, T > C, G > A, G > T etc) can be deduced from the entries in the body of the table against the Hu-1 line at the top of each VSD figure (NC_045512). Note that a detailed compilation of the main types of SNV changes at APOBEC deamination C-sites and ADAR deamination A-sites for isolates Jan-Mar 2020 from Northern Hemisphere hotspot zones (Wuhan, New York City, West Coast USA, Spain) can be found in Steele and Lindley (2020).¹⁴

3.3.1 | March

The VSD in Supporting Information File D show the sequence patterns of SNVs across the SARS-CoV-2 full-length L241f.1 genomes collected in March 2020. Note the striking sequence conservation at the RNA and at the amino acid level. In this sample, 12/33 [36%] are unmutated and many are lightly mutated (15/33 [46%]) with 1-2 SNV per sequence. The sharing of sequences in the lightly mutated category can be seen. P-to-P transfers may well be evident in the unmutated set, but patient metadata would be required to establish that. These sequence data suggest that patients from SARS-CoV-2-infected Northern Hemisphere zones brought the unmutated L241f.1 sequence into Melbourne, and if they did transfer the sequence on by secondary P-to-P the recipient subject on average introduced one or two new SNVs into the sequence. This was also observed repeatedly in our first report.¹⁴ The smaller numbers of more heavily mutated sequences (eg MT451254 with 8 SNVs) may indicate the endpoint of possibly two P-to-P transfers, and we assume in healthy subjects with active APOBEC and ADAR deaminase mutator responses. In the same vein, some of the unmutated sequences could be primary or secondary P-to-P infections circulating in subjects with deficits in Innate Immune deaminase mutagenesis.

3.3.2 | April through May

These sample VSD alignments (Supporting Information D) capture the VSD patterns from March, April through to May 19, and reveal the emergence of the L241f.1vic sequence in collections on May 13 (unmutated or lightly

mutated). There are other interesting features in these data including sharing of sequences indicative of a number of P-to-P secondary transfers to at least 1 or 4 other recipient patients (close interacting groups such as a family?). Notice clear additional mutagenesis on secondary transfer in these groups, suggesting these recipients could be healthy subjects with functioning Innate Immune deaminase mutagenesis. The L241f.1vicm sequence identified did not flourish much further during June and early July (Supporting Information File D).

3.3.3 | June 5–June 25

This sample alignment (Supporting Information File D) illustrates features already discussed and shows that the L241f.1vic sequence begins to dominate the sequences uploaded to NCBI Virus through June. Note that a residue of L241f.1 sequences, initiated as secondary transfers and thus derivatives from incoming overseas travellers (?) display sequences that are heavily mutated, in contrast to many other sets (both L241f.1vicm and L241f.1vic sequences). What is notable about the L241f.1vicm and L241f.1vic sequences is largely the absence of mutation, or evidence of lightly mutated sequences on P-to-P transfer (Table 3).

3.3.4 | June 23–June 30

These sample alignments (Supporting Information D) show a continuation of the trend discussed (Table 3). Thus, largely unmutated representatives of the L241f.1vic haplotype clearly dominate all uploaded sequence collections from infected subjects. Again, on secondary transfer of the unmutated sequence, one or two SNVs may be introduced but most are not further mutated or lightly mutated (Table 3).

3.3.5 | June 28–July 22

The alignments over these collection periods (Supporting Information File D) were for samples collected when the explosive epidemic was reaching its peak (Figure 1). Whilst there are still many unmutated and lightly mutated L241f.1vic sequences (1-2 SNVs per sequence), there is also evidence of multiple P-to-P secondary transfers with a higher number of SNVs per sequence (Table 3)—this could plausibly suggest transfers in closed aged care facilities to healthier medical and carer staff (or other visitors) in such institutions. This latter group dominate sets

of cases, outnumbering infected residents by about fifty per cent (Supporting Information File F).

3.3.6 | July 28 to August 23

These alignments (Supporting Information File D) cover sequence collections as the epidemic was declining though August into September (right hand side of Figure 1). The increased SNV density pattern per sequence has now become far more pronounced. Whilst there are clearly some unmutated L241f.1vic sequences, the great majority are now mutated and evidence of secondary transfer of group sizes 2-4 are now evident in the data (Table 3). The dominance of clearly multiply mutated sequences is plausible evidence of P-to-P transmission into and from healthy subjects introducing Innate Immune deaminase-mediated mutations into the viral genome (which could also be introductions, or infections, into elderly co-morbid patients as well, but patient metadata would be required to identify such P-to-P transfers).

3.3.7 | July 23 to September 8

These alignments (Supporting Information File D) cover sequence collections as the epidemic was in evident decline though August into September (right-hand side of Figure 1). The patterns of SNV and P-to-P transfers are very similar to that discussed above in Section 3.3.6 July 28 to August 23 (Table 3).

The data just presented is either comprehensive or were selected to be representative and illustrative of the SARS-CoV-2 sequence changes in Victoria in the 1st and 2nd Waves (as made public at the NCBI Virus website). The wider set of SNV patterns showing this clear quantitative and qualitative trend in the VSD plots can be found in Supporting Information File D. The clear trend from mid-June to end of July is the amplification of the L241f.1vic haplotype clone virtually unmutated (or very lightly mutated) as summarized also in Table 3.

3.4 | Biases in uploaded sequence data

The uploaded sequence data at NCBI Virus for collections in Victoria, Australia, whilst substantial are clearly biased and do not report the complete numbers of known full-length genomes, estimated to be approximately 3500-4000 'Community Transmissions' or 'Mystery Cases' described above. For the period covering the entire 2nd Wave June 1-September 30, the numbers of these acquired cases from an unknown source

totalled 3571 (from data in Figures 3-5). That is, we do not know the extent of the types of haplotype diversity (see Table 2) or mutational loads in such haplotypes of such subjects, nor do we have information on such 'Mystery Cases' amongst residents and other cases in aged care facilities (see Supporting Information F). The clear biased deficit in numbers of putative mystery case SARS-CoV-2 genomic sequences, which would have differed significantly in haplotype sequence from the main amplified L241f.1vic haplotype (unmutated or lightly mutated), requires further comment.

Either these collections were not sequenced: that is, contact tracing established by conventional means failed to identify a confirmed case contact X was considered sufficient evidence. Alternatively, they were routinely sequenced given the scale of the sequencing operation at The Peter Doherty Institute. This seems more plausible as all public commentary through June, July and August including testimony by lead epidemiologist at the Victorian Department of Health (Dr Charles Alpre) to the formal government inquiry into aged care outbreaks²⁷ directly implied that mystery cases were full-length genomes sequenced to formally exclude them from the main known hot spot clusters in aged care. They were not made publicly available to GenBank/NCBI Virus. We can only speculate that they were not made available as they were considered 'uninteresting sequences' in explaining the wild-fire spread of the main SARS-CoV-2 'clone' (L241f.1vic) through Victorian aged care and nursing facilities. This is a possible yet speculative explanation given the medical emergency at the time. Indeed as reported in the *British Medical Journal*, Professor Ben Howden, director of the Microbiological Diagnostic Unit Public Health Laboratory at The Peter Doherty Institute, told the government inquiry²⁹ in mid-August 2020.. '...that, of the 1837 cases of local transmission that have been sequenced since 8 May, 99.8% came from three clusters, one from the Rydges Hotel and two from the five star Stamford Plaza Hotel'.²⁹ We can plausibly associate the major L241f.1vic haplotype with the Rydges Hotel cluster, and the other minor haplotypes through May and June with the L241f.1vicm, L241a, L241c and L241f.1 haplotypic variants (Figure 6, Supporting Information Files B, D and section e. below, analysis of GISAID db available sequencing data). Yet everyday throughout July and August official public references were made (by the Premier of Victoria Hon. Daniel Andrews and Chief Health Officer Professor Brett Sutton) to the continuing high numbers of mystery cases as it was feared they were contributing also by allowing transmissions to aged care facilities causing infection cluster flare ups (the spikes in daily case numbers as seen in Figure 1).

TABLE 4 Total statistics in Victorian aged care in 2nd wave 2020

Confirmed cases	Staff cases	Resident cases	Other cases	Deaths
5078	2009	1947	1122	647
-	(39.6%)	(38.3%)	(22.1%)	(12.7%)

Note: Summary of totals from Supporting Information F.

3.5 | Haplotype analyses of GISAID publicly available genomic sequencing data

An independent check at the GISAID db (and GenBank db) on total Victorian sequences and their haplotypes (Table 2) was conducted by Dr Jared Mamrot on the sequences uploaded by The Peter Doherty Institute for collections in the period June 1 through September 30. The magnitude of sequence numbers established in the analysis at NCBI Virus and the dominance of the L241f.1vic sequences in June, July, August and September were confirmed. In addition, the sporadic occurrences of other minor haplotypes are consistent with our above analyses, in June 2020. We have concluded that genomic sequences from the cases from an unknown cause, that is the expected large numbers of sequences quite different from the dominant L241f.1vic sequence, were not present in the GISAID repository at the time of submission of this paper. The haplotyping analysis is in Supporting Information File G. In this separate analysis, all Victorian GISAID and associated GenBank ID sequences (at both low and high coverage) allow estimates of the extreme haplotype dominance of the L241f.1vic haplotype (>99% of all sequences collected) and the occurrence of other minor variants and haplotypes: the 18 nt in-frame deletion variant of L241f.1vic, from p.288879 removing the p.28881-3 marker (Supplementary Information File B); the minor L241f.1vicm variant ($n \sim 214$) detected in May and June in GenBank/NCBI Virus uploaded data, which did not further appear in high numbers through July and August in sequence data uploaded to both GenBank/NCBI Virus and GISAID databases; and lower numbers of L241f $n \sim 5$, L241fc $n \sim 41$, L/Ln $n \sim 6$ and S with $n \sim 1$.

4 | DISCUSSION

In our considered view, as listed above, there are a number of scientific advantages in analysing the data on what exactly happened at the SARS-CoV-2 genomic sequence level during the 1st and 2nd Waves in Victoria, Australia. Thus, despite the extreme personal hardships, the SARS-CoV-2 epidemics in Australia enabled many types of observations based on a set of almost 'controlled conditions'

or 'biomedical experiments of nature' (despite the biases in the data highlighted). For example, as we discuss below, it allows a considered reflection on how the 'UK Mutant' first arose to dominance and why this publicized highly putatively contagious SARS-CoV-2 mutant did not spread when it appeared in Australia in January 2021 brought in by multiple travellers from the UK into many different major Australian cities (Adelaide, Perth, Brisbane, Melbourne, Sydney).

Using publicly available data and observations, we have recorded here some key genomic features of the 1st and 2nd Wave SARS-CoV-2 epidemics in Victoria that can be discerned which involves mainly the city of Melbourne and some nearby regional towns and cities to the north west (Macedon, Mitchell, Bendigo, Ballarat) and south-west of the city along Port Phillip Bay (Geelong, Colac) comprising upwards of 5 million people in a small area of land mass (as the greater eastern part of the state was basically virus free). The large numbers of publicly available SARS-CoV-2 genomes at the NCBI Virus database could be systematically analysed by the simple accessible alignment tools provided at that site. This has allowed us considerable insight into the mode of SARS-CoV-2 haplotype variation strategies and the major infection transmission mode of the virus in the Victorian aged care and elderly community.

To summarize our main findings.

- a. The severity and quantitative patient dominance of the Victorian 2nd Wave epidemic in 2020 was focussed almost exclusively on elderly citizens (Tables 1 and 4), presumed vulnerable and immune defenceless as discussed already due to clear deficits in first-line IFN-driven innate immunity.^{4,11} Many of these 'immune defenceless' elderly co-morbid citizens were concentrated in closed aged care communities (Supporting Information F and Table 4). Given that APOBEC and ADAR innate immune 'mutators' are activated by the same IFN type I and type III stimulated gene pathways¹⁹⁻²⁶ it is a logical priority to consider the level of their mutagenic impact on the SARS-CoV-2 clones being transferred within and between aged vulnerable groups (eg VSD analysis in the Supporting Information D).
- b. It is clear that the L241f.1vic clone dominated the effective 'fire storm' of infections in the Victorian elderly. The spread of *largely unmutated clones* of this haplotype, particularly in second half June 2020 into late July 2020, is striking and suggests little or no APOBEC and ADAR mutagenesis on P-to-P transfers, and this plausibly implies the infections were rampant in patients with clear deficits in IFN type I and type III stimulated gene pathways (see in particular the VSD patterns in

Supplementary Information Files D and F, and Table 3 summary).

The data reported herein are thus consistent with the following P-to-P infection model which is also the operational hypothesis under test: clusters of immune defenceless elderly co-morbid citizens in many aged care and nursing facilities were all systematically struck with devastating force (high infection rates and death rates), with a single unmutated (or lightly mutated) SARS-CoV-2 haplotype variant (L241f.1vic). Through late June, July, August and September in 2020, this putatively cloned variant must have been spread unimpeded by carriers who were asymptomatic or lightly symptomatic infected health-care professionals and carers working across multiple age care institutions.³⁰⁻³³ The large-scale amplifications of the L241f.1vic variant—instanced by the size of the multiple 'New case' spikes (shown in Figure 1), particularly through July 2020—could have produced many trillions of L241f.1vic virions in each location thus contaminating numerous surfaces (fomites, personal effects of all types) and could have contaminated or infected human carriers in each institution. This then fuelled the further putative quantitative dominant rapid spread of this apparently capricious L241f.1vic variant into the local community and particularly to other aged care facilities leading to further putative viral amplifications in elderly co-morbid subjects. If anything, the Victorian experience underlines why elderly co-morbid citizens require very special care, protection and therapies during cold and flu seasons.^{31,32} Such an approach now seems obvious to handle the viral-induced respiratory crises in vulnerable aged care patients. There is a need for much further research on how to boost innate and specific adaptive immunity defences in such vulnerable patients—so as to achieve a survival infection outcomes in such vulnerable patients (see Sette and Crotty¹⁰ integrated summary, and Lucas et al¹¹).

The fact that the L241f.1vic variant remained largely unmutated or very lightly mutated on secondary transfer (1-2 SNVs, Table 3) implies that the main amplifying source of viral replication in each aged care facility were patients with extreme deficits in first-line innate immune IFN type I and III dependent anti-viral immunity, see Figure 2C in Lucas et al¹¹. The levels of both APOBEC and ADAR host deaminase anti-viral mutators are normally induced by invading pathogens in a healthy cellular innate immune response.¹⁹⁻²⁵ The lack of apparent mutation in the main transmitted L241f.1vic clone through June and July is consistent with the hypothesis that the virus was replicating almost unimpeded in immune defenceless elderly co-morbids. Indeed many travellers into Australia carrying the L241f.1 haplotype variant in March-April 2020, and their presumed putative secondary

transfer recipients also display unmutated or lightly mutated sequences (Supporting Information File D). This suggests those incoming overseas travellers were probably also elderly immune defenceless patients. There is further support for this interpretation. Through August and into September, the incidence and frequency of mutations in mutated derivatives of the L241f.1vic haplotype increased, significantly, a pattern consistent with P-to-P transfer and passage of the L241f.1vic clone through healthy asymptomatic carriers with intact innate immune system and functioning APOBEC/ADAR mutators (Table 3, Supporting Information File D).

Our further hypothesis is that the L241f.1vic variant, which emerged as an escape haplotype variant from a hotel quarantine site on May 8 (Rydges Hotel)²⁷ gained a low level P-to-P foothold in the community for a month through late May and into mid-June where upon it was fortuitously and explosively amplified and spread on scale when introduced into closed groups of immune defenceless elderly co-morbid patients living in the Kensington Towers in north west Melbourne and aged care facilities.²⁹ Through late June and into July, its large amplification factors and numerous cross-institutional spreading ‘vectors’ (asymptomatic family members, aged carers and health workers, and their presumed fomites) ensured it took over and spread quickly through all of Melbourne’s and nearby regional aged care facilities. The sequencing effort by the Victorian Department of Health in association with The Peter Doherty Institute and epidemiological interpretation was thus focussed on the L241f.1vic haplotype to try and prevent its further spread.

Our further related hypothesis is that the much publicized ‘UK Mutant’ (B.1.1.7) that emerged in the United Kingdom in the latter months of 2020 spread and came to dominance in much the same way as L241f.1vic in Victoria, Australia—a capriciously amplified SARS-CoV-2 variant in closed communities of immune defenceless elderly co-morbid patients.

The UK Mutant is of the L241f.1 haplotype (L241f.1uk)—but different from L241f.1vic which has a smaller number ($n = 7$) of ‘riboswitch’ type key distinctive SNV markers that distinguish it from L241f.1 (Table 2)—with about 17–19 additional sequence changes from L241f.1, four being deletions and one a STOP codon (Rambaut et al¹⁶). Apart from distinctive radicle changes, all the other SNV changes in L241f.1uk appear conserved or benign, a picture consistent with all our sequence analysis here and in Steele and Lindley.¹⁴ The advance news media notice on the ‘UK Mutant’ as it arrived with travellers into Australia in early 2021 described the variant as a ‘highly contagious rapidly spreading mutant’. On arrival with travellers from overseas, it inevitably escaped quarantine into the community

and in public airport domestic departure lounges in Adelaide, Brisbane, Perth and then Melbourne exposing potentially many hundreds of Australian residents and other travellers. These escapes each caused severe hard Stage 4 lock downs (2–5 days) of the entire city populations, and in Victoria and Western Australia almost the entire State populations for 5 days. Given all the advance publicity and the hundreds of putative reported P-to-P contacts and opportunities for further P-to-P spreading (and further interstate by aeroplanes by many more hundreds of putative contacts), the UK Mutant was *not contagious* and did not spread into the wider Australian community.^{34,35} These very clear negative transmission data led the noted Australian epidemiologist at The Australian National University, Professor Peter Collignon, to release his considered opinion to *The Australian* newspaper³⁴:

‘...In retrospect, the Melbourne lockdown was unnecessary... From my perspective, if you’ve got very little community transmission, I’m not sure that a short lockdown achieves much extra, if you’ve got good contact tracing and good testing’, he said. ... ‘If I look at the lockdowns done in Adelaide, Brisbane, Perth, and now Melbourne, it didn’t turn up one more case than contact tracing did. ... ‘The UK strain has not spread uncontrollably and wildly’.

Clearly, further research is required to understand exactly why the ‘UK Mutant’ did not spread when multiply introduced via multiple entry points into Australia in the first few months of 2021. One possibility is that it really was not that contagious in normal healthy people, it had been attenuated but behaved as a putative capriciously amplified variant in elderly co-morbids. That is clearly dangerous for that vulnerable group (which by chance were not in large numbers in the public places the infected travellers were supposed to have visited).

In our view, what has unfolded with the L241f.1vic, as documented here, is a clear putative exemplar. Our hypothesis is that the ‘UK Mutant’ (B.1.1.7) or L241f.1uk haplotype was a capricious P-to-P generated SARS-CoV-2 variant haplotype that was fortuitously amplified on scale to quantitative clonal dominance in and between aged care facilities in southern England from September 2020, much like the L241f.1vic haplotype in Victoria, Australia. The main difference between the otherwise similar countries (both economically, culturally and multi-ethnic diverse composition) is one of population scale given the population differential between UK and Victoria is about ten-fold, and given similar land areas the population density in the UK would be far higher than Victoria.

Given all these findings what then of appropriate vaccine design? Apart from any safety or adverse reaction issues with the current set of spike protein mRNA expression vector vaccines, there are, in our opinion, three main issues to be addressed, two arise from our analysis. These relate in general to the vulnerable *target group* to be protected—the immune defenceless elderly co-morbid—and *the type* of protective acquired immunity to be induced by any vaccine (the relevant vaccine data is under detailed review elsewhere Gorczynski RM, Lindley RA, Steele EJ, Wickramasinghe NC. 2021 ‘Nature of acquired immune responses, epitope specificity and resultant protection from SARS-CoV-2’, under submission).

First, boosting innate immunity would seem a priority medical strategy in the vulnerable elderly, as for example BCG vaccination is expected to do for the lung infection tuberculosis. This could involve the use of IFNs to stimulate deaminase expression during the acute stage of infection. BCG is a safe and widely accepted non-specific activator of innate immunity, and it could contribute to elevated APOBEC and ADAR expression and thus mutagenesis of the coronavirus genome. This logically implies that strategies for boosting or elevation, by ‘training’ the innate immune responses^{7,12} may boost mutagenic APOBEC and ADAR expression levels and other anti-viral gene products.^{19,20}

Second, it is still unclear whether the current crop of SARS-CoV-2 vaccines is inducing *the appropriate class of acquired immunity*, namely mucosal responses involving specific secretory IgA responses in nose, saliva and respiratory tract (A. Hapel, discussion and person comm. February 2021). This is also confirmed by the April 20 2021 update report on vaccine efficacy from the US Centers for Disease Control.³⁶ A careful reading of that report suggests that none of the current ‘jab in the arm’ vaccines protect against catching SARS-CoV-2, yet may moderate severity. This is consistent with prior historical immunology discoveries. It has been known for over 45 years that the best form of protective immunity for pathogens invading by the nasal or oral route are local secretory IgA responses.^{37,38} Therefore, at this juncture, it is not clear whether currently deployed intramuscular (im) parenteral immunization is the best approach for highly effective mucosal immunity against SARS-CoV-2, and the oral-nasal routes of vaccination should be considered. Indeed, recent informed analysis and commentary supports this conclusion.^{39,40} Natural infections with SARS-CoV-2 (in recovered patients) would therefore be expected to induce protective dimeric sIgA mucosal immunity. Certainly, the recent longitudinal population scale study in Denmark implies that prior infection with SARS-CoV-2 affords upwards of 80% protection in the population under 65 years against reinfection between the first and second

major surges of SARS-CoV-2 in Denmark in 2020; with the protective rate in the re-infected elderly vulnerable group a half lower again at 47%.¹³ These are encouraging findings suggesting, at the time of writing, that natural ‘herd immunity’ could be well underway in Denmark and similar Northern hemisphere infected zones in 2020 and into likely surges and waves of SARS-CoV-2 in 2021.

Finally, a striking feature of the adaptation strategy of the SARS-CoV-2 coronavirus is the signature of ‘haplotype ribo-switching’, focussed mainly on C-site and A-site centred motifs targeted by APOBEC and ADAR deamination events.¹⁴ We expect this strategy to be general for coronavirus infections (it is evident in the additional SNVs over the L241f.1 haplotype in the L241f.1vic variant). Thus, whilst some variation can occur at the amino acid level, it is largely conserved as discussed¹⁴ and as illustrated in the haplotyping data of the VSD presented here (Supporting Information File D). The predominant conclusion is the large amount of viral variation that goes on at the RNA level with presumptive links to functional RNA secondary structures, that is downstream consequences for appropriate RNA secondary structures compatible with viral replicative fitness. The striking sequence conservation at the amino acid level after human passage¹⁴ implies that there are key conserved variations mainly at the RNA genomic level. Thus, most recorded SNVs were either synonymous or, if non-synonymous, the SNVs in coding regions lead often to likely functionally benign amino acid outcomes and thus result in likely conservation of viral function at the protein level. This was a striking fact across all data sets examined¹⁴ and in the data reported here (Supporting Information File D). Clearly, all these observed viral sequences survived Darwinian negative selection within each infected subject against malfunctioning and poorly replicating variants in the quasi-species pool as discussed.¹⁴ Targeted mutagenesis studies^{41,42} on viral genomes affirm this conclusion. That data demonstrate how sensitive and tightly tuned information-rich single-stranded viral genomes really are to *any* non-synonymous substitutions of many types. In Hepatitis C Virus (HCV) studies, another similar approach has demonstrated how targeted substitutions in the coding regions of the HCV genome reveals a functional network of RNA regulatory structures important in efficacy of viral replication and infectivity OK.⁴³

Therefore, new vaccine designs and therapies may need to target these new and putative stable secondary structures thrown up by haplotype ribo-switching for the presumptive replicative stability of the viral RNA genome. This type of RNA secondary structure research and analysis is underway (Lindley RA and associates). This will obviously involve an entirely different approach to vaccine RNA target site design and curative therapies, which will necessitate further research on RNA structure rather than

the protein components of the virus (aspects reviewed in Robson et al¹⁸).

ACKNOWLEDGMENT

We thank Dr Jared Mamrot of GMDxgen Pty Ltd for the independent analysis of the haplotype sequences of Victoria Australia genomes made publicly available at the GISAID database. We thank Dr Andrew Hapel and Professor Patrick Carnegie for discussions and references on secretory IgA mucosal immunity.

CONFLICT OF INTEREST

There is no conflict of interest with the scientific objectivity of this paper.

AUTHOR CONTRIBUTIONS

RAL and EJS conceived the plan, aim and organization of the paper, and EJS wrote the first draft of the paper. EJS did the primary analysis of GenBank/NCBI Virus accessed COVID-19 genomic sequences. Both authors read and worked on the final version of the paper before submission by EJS.

ORCID

Robyn A. Lindley  <https://orcid.org/0000-0002-8952-1536>

Edward J. Steele  <https://orcid.org/0000-0003-4700-3156>

REFERENCES

- Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *Lancet*. 2020;395:497-506.
- Cohen J. Wuhan seafood market may not be source of novel virus spreading globally. *Science*. 2020;367. <https://www.sciencemag.org/news/2020/01/wuhan-seafood-market-may-not-be-source-novel-virus-spreading-globally>
- Pekar J, Worobey M, Moshiri N, Scheffler K, Wertheim JO. Timing the SARS-CoV-2 index case in Hubei province. *Science*. 2021;372:412-417. doi:10.1126/science.abf8003
- Acharya D, Liu G-Q, Gack MU. Dysregulation of type I interferon responses in COVID-19. *Nat. Rev Immunol*. 2020;20:397-398. doi:10.1038/s41577-020-0346-x
- Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. 2020;181:1036-1045. doi:10.1016/j.cell.2020.04.026
- Hadjadj J, Yatim N, Barnabei L, et al. (2020) Impaired type I interferon activity and exacerbated inflammatory responses in severe Covid-19 patients medRxiv preprint doi:10.1101/2020.04.19.20068015
- Netea MG, Giamarellos-Bourboulis EJ, Dominguez-Andre's J, et al. Trained Immunity: A tool for reducing susceptibility to and the severity of SARS-CoV-2 infection. *Cell*. 2020;181:969-977.
- Zhang Q, Bastard P, Liu Z, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. 2020;370:eabd4570.
- Moderbacher CR, Ramirez SI, Dan JM, et al. Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell*. 2020;183:996-1012.
- Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell*. 2021;184:1-20.
- Lucas C, Wong P, Klein J, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*. 2020;584:463-469.
- Netea MG, Dominguez-Andres J, Barreiro LB, et al. Defining trained immunity and its role in health and disease. *Nat Rev Immunol*. 2020;20:375-388.
- Hansen CH, Michlmayr D, Gubbels SM, Mølbak K, Ethelberg S. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: A population-level observational study. *The Lancet*. 2021;397 doi:10.1016/S0140-6736(21)00575-4
- Steele EJ, Lindley RA. Analysis of APOBEC and ADAR deaminase-driven Riboswitch Haplotypes in COVID-19 RNA strain variants and the implications for vaccine design. *Res Rep*. doi:10.9777/rr.2020.10001 <https://companyofscientists.com/index.php/rr>
- Rambaut A, Holmes EC, Hill V, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *Nat. Microbiol*. 2020;5:1403-1407.
- Rambaut A, Loman N, Pybus O, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virologica.org In Press* Feb 4 2021 <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>
- Forster P, Forster L, Renfrew C, Forster MJ 2020 Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci USA*. 2020;117:9241-9243.
- Robson F, Khan KS, Le TK, et al. Coronavirus RNA proof-reading: Molecular basis and therapeutic targeting. *Mol Cell*. 2020;79:710-727.
- Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol*. 2011;1:519-525.
- Schneider WM, Chevillotte MD, Rice CM. Interferon-stimulated genes: A complex web of host defenses. *Annu Rev Immunol*. 2014;232:513-545.
- Samuel CE. Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology*. 2011;411:180-193. doi:10.1016/j.virol.2010.12.004
- Coticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol*. 2008;9:229. doi:10.1186/gb-2008-9-6-229
- Harris RS, Bishop KN, Sheehy AM, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell*. 2003;113:803-809.
- Refsland EW, Harris RS. The APOBEC3 family of retroelement restriction factors. *Curr Top Microbiol Immunol*. 2013;371:1-27. doi:10.1007/978-3-642-37765-5_1
- Vartanian J-P, Henry M, Marchio A, et al. Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog*. 2010;6(5):e1000928. doi:10.1371/journal.ppat.1000928

26. Lindley RA. Review of the mutational role of deaminases and the generation of a cognate molecular model to explain cancer mutation spectra. *Med Res Arch*. 2020;8(8):2177.
27. COVID-19 Hotel Quarantine Inquiry, Final Report and Recommendations, Volume I Parl paper no. 191 (2018–2020) ISBN 978-0-6450016-1-7 (Victorian Government Inquiry).
28. Seemann T, Lane CR, Sherry NL, et al. Tracking the COVID-19 pandemic in Australia using genomics. *Nat Com*. 2020;11(1):4376. doi:10.1038/s41467-020-18314-x
29. Smith P. Covid-19 in Australia: most infected health workers in Victoria's second wave acquired virus at work. *BMJ*. 2020;370:m3350. doi:10.1136/bmj.m3350
30. ABC News. "Workplace coronavirus transmission driving Victorian case numbers, including in aged care" Sunday 19 July 2020 at 2:27pm, updated Monday 20 July 2020 at 12:30 am. <https://www.abc.net.au/news/2020-07-19/workplace-coronavirus-transmission-in-victoria-in-aged-care/12470704>
31. Cousins S. Experts criticise Australia's aged care failings over COVID-19. *The Lancet*. 2020;396:1322-1323.
32. Hashan MR, Smoll N, King C, et al. Epidemiology and clinical features of COVID-19 outbreaks in aged care facilities: A systematic review and meta-analysis. *EClinicalMedicine*. 2021;33:100771. doi: 10.1016/j.eclinm.2021.100771
33. Rose T Carer tracking lacking. Herald-Sun, Melbourne 31 August 2020 p.12
34. Baxendale R, Robinson N. Spread of UK coronavirus variant limited to close contacts. *The Australian*. February 20, 2021
35. Robinson N. Why didn't others get the hotel virus. *The Australian* February 9, 2021.
36. CDC Update: Science Brief: Background Rationale and Evidence for Public Health Recommendations for Fully Vaccinated People Centers for Disease Control and Prevention, COVID-19 Updated Apr. 2, 2021 <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html>
37. Steele EJ, Chaicumpa W, Rowley D. Isolation and biological properties of three classes of rabbit antibody to *Vibrio Cholerae*. *J Infect Dis*. 1974;130:93-103.
38. Steele EJ. 1975. Efficiency of antibody classes in cholera immunity. PhD diss. University of Adelaide.
39. Bleier BS, Ramanathan M, Lane AP. COVID-19 vaccines may not prevent nasal SARS-CoV-2 infection and asymptomatic transmission. *Otolaryngol Head Neck Surg*. 2021;164:305-307. doi:10.1177/0194599820982633
40. Schiavone M, Gasperetti A, Mitacchione G, Viecca M, Forleo GB. Response to: COVID-19 re-infection vaccinated individuals as a potential source of transmission. *Eur J Clin Invest*. 2021;00:e13544. doi:10.1111/eci.13544
41. Sanjuan R, Moya A, Elena SF. The distribution of fitness effects caused by single- nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci USA*. 2004;2004(101):8396-8401.
42. Sanjuan R, Nebot MR, Chirico N, et al. Viral mutation rates. *Virology*. 2010;84:9733-9748.
43. Pirakitikulr N, Kohlway A, Lindenbach BD, Pyle AM. The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol Cell*. 2016;61:1-10. doi:10.1016/j.molcel.2016.01.024

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Lindley RA, Steele EJ. Analysis of SARS-CoV-2 haplotypes and genomic sequences during 2020 in Victoria, Australia, in the context of putative deficits in innate immune deaminase anti-viral responses. *Scand J Immunol*. 2021;94:e13100. <https://doi.org/10.1111/sji.13100>