Virus Evolution, 2020, 6(1): veaa027

doi: 10.1093/ve/veaa027
Rapid Communication

An emergent clade of SARS-CoV-2 linked to returned travellers from Iran

John-Sebastian Eden,[1,2,*,†] Rebecca Rockett,[1,3,4] Ian Carter,[3] Hossinur Rahman,[3] Joep de Ligt,[5] James Hadfield,[6] Matthew Storey,[5] Xiaoyun Ren,[5] Rachel Tulloch,[1,2] Kerri Basile,[3*] Jessica Wells,[3] Roy Byun,[7] Nicky Gilroy,[3] Matthew V. O'Sullivan,[3,4] Vitali Sintchenko,[1,3,4] Sharon C. Chen,[1,3,4] Susan Maddocks,[3] Tania C. Sorrell,[1,2,3] Edward C. Holmes,[1,‡] Dominic E. Dwyer,[1,3,4] and Jen Kok[3,4]; for the 2019-nCoV Study Group[§]

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences & School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia, [2]Centre for Virus Research & Centre for Infectious Diseases and Microbiology, Westmead Institute for Medical Research, PO Box 412, Westmead, NSW 2145, Australia, [3]Centre for Infectious Diseases and Microbiology Laboratory Services, NSW Health Pathology – Institute of Clinical Pathology and Medical Research, Westmead, NSW 2145, Australia, [4]Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital, Westmead, NSW 2145, Australia, [5]Institute of Environmental Science and Research, Porirua 5240, New Zealand, [6]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA and and [7]NSW Ministry of Health, North Sydney, NSW 2059, Australia

*Corresponding authors: E-mails: js.eden@sydney.edu.au; kerri.basile@health.nsw.gov.au

[†]https://orcid.org/0000-0003-1374-3551

[‡]https://orcid.org/0000-0001-9596-3552

[§]The members of 2019-nCoV Study Group are listed in the Acknowledgments.

Abstract

The SARS-CoV-2 epidemic has rapidly spread outside China with major outbreaks occurring in Italy, South Korea, and Iran. Phylogenetic analyses of whole-genome sequencing data identified a distinct SARS-CoV-2 clade linked to travellers returning from Iran to Australia and New Zealand. This study highlights potential viral diversity driving the epidemic in Iran, and underscores the power of rapid genome sequencing and public data sharing to improve the detection and management of emerging infectious diseases.

Key words: COVID-19; SARS-CoV-2; genome sequencing; phylogenetics.

© The Author(s) 2020. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

1. Introduction

From a public health perspective, the real-time whole-genome sequencing (WGS) of emerging viruses enables the informed development and design of molecular diagnostic assays, and tracing patterns of spread across multiple epidemiological scales (i.e. genomic epidemiology). However, WGS capacities and data sharing policies vary in different countries and jurisdictions, leading to potential sampling bias due to delayed or underrepresented sequencing data from some areas with substantial SARS-CoV-2 activity. Herein, we show that the genomic analyses of SARS-CoV-2 strains from Australian returned travellers with COVID-19 disease may provide important insights into viral diversity present in regions currently lacking genomic data.

2. SARS-CoV-2 emergence and dissemination

In late December 2019, a cluster of cases of pneumonia of unknown aetiology in Wuhan city, Hubei province, China was reported by health authorities (Wuhan Municipal Health Commission 2019). A novel betacoronavirus, designated SARS-CoV-2, was identified as the causative agent (Wu et al. 2020) of the disease now known as COVID-19, with substantial human-to-human transmission (Lu et al. 2020). To contain a growing epidemic, Chinese authorities implemented strict quarantine measures in Wuhan and surrounding areas in Hubei province. Significant delays in the global spread of the virus were achieved, but despite these measures, cases were exported to other countries. As of 9 March 2020, these numbered more than 100 countries, on all continents except Antarctica; the total number of confirmed infections exceeded 110,000 and there were nearly 4,000 deaths (Dong, Du, and Gardner 2020). Although the majority of cases have occurred in China, major outbreaks have also been reported in Italy, South Korea, and Iran (World Health Organisation 2020a). Importantly, there is widespread local transmission in multiple countries outside China following independent importations of infection from visitors and returned travellers.

3. WGS of SARS-CoV-2 cases in Australia and New Zealand

Viral extracts were prepared from respiratory tract samples where SARS-CoV-2 was detected by reverse-transcription polymerase chain reaction (RT-PCR) using World Health Organisation (2020b) recommended primers and probes targeting the E and RdRp genes. In New South Wales (NSW), Australia, WGS for SARS-CoV-2 was developed based on an existing amplicon-based Illumina sequencing approach (Di Giallonardo et al. 2018). Viral extracts were reverse transcribed with SSIV VILO cDNA master mix and then used as input for multiple overlapping PCR reactions (~2.5 kb each) spanning the viral genome using Platinum SuperFi master mix (primers provided in Supplementary Table S1). Amplicons were pooled equally, purified, and quantified. Nextera XT libraries were prepared and sequencing was performed with multiplexing on an Illumina iSeq (300 cycle flow cell). In New Zealand, the ARTIC network protocol was used for WGS (Quick 2020). In short, 400-bp tiling amplicons designed with Primal Scheme (Grubaugh et al. 2019a) were used to amplify viral cDNA prepared with SuperScript III. A sequence library was then constructed using the Oxford NanoPore ligation sequencing kit and sequenced on a R9.4.1 MinION flow cell. Near-complete viral genomes were then assembled *de novo* in Geneious Prime 2020.0.5 or through reference mapping with RAMPART V1.0.6 (Hadfield 2019) using the ARTIC network nCoV-2019 novel coronavirus bioinformatics protocol (Loman and Rambaut 2020). In total, 13 SARS-CoV-2 genomes were sequenced from cases in NSW diagnosed between 24 January and 3 March 2020, as well as a single genome from the first patient in Auckland, New Zealand sampled on 27 February 2020 (Table 1). Australian and New Zealand sequences were aligned to global reference strains sourced from GISAID with MAFFT (Katoh 2002) and then compared phylogenetically using a maximum-likelihood approach—PhyML v2.2.4 (Guindon and Gascuel 2003).

4. A distinct clade of SARS-CoV-2 identified in travellers returned from Iran

The Australian strains of SARS-CoV-2 were dispersed across the global SARS-CoV-2 phylogeny (Fig. 1A). The first four cases of COVID-19 disease in NSW occurred between 24 and 26 January 2020, and these were closely related with 1–2 single nucleotide polymorphisms (SNPs) difference to the dominant variant circulating in Wuhan at the time (prototype strain MN908947/SARS-CoV-2/Wuhan-Hu-1). As the four patients identified in this period had recently returned from China, this region was the likely source of infection. From 1 February 2020, travel to Australia from mainland China was restricted to returning Australian residents and their children, who were placed in home quarantine for 14 days. Despite the intensive testing of such returning travellers, no further cases of COVID-19 were detected in NSW until 28 February 2020, when SARS-COV-2 was detected in an individual returning from Iran (NSW05). A close contact of this individual also tested positive (NSW14) providing the first evidence of local transmission within NSW. This was followed by further Iran travel-linked cases in NSW (NSW06, NSW11, NSW12, and NSW13) and New Zealand (NZ01).

Of note, the genomes of all patients with a history of travel to Iran were part of a monophyletic group defined by three nucleotide substitutions (G1397A, T28688C, and G29742T) in the SARS-CoV-2 genome relative to the Wuhan prototype strain (Fig. 1B). G1397A and T28688C both occur in coding regions with G1397A producing a non-synonymous change (V378I) in the ORF1ab-encoded non-structural protein 2 region. G29742T occurs in the 3′-UTR. In addition to the Australian and New Zealand strains, this clade also included a traveller who had returned to Canada from Iran (BC_37_0-2), providing further evidence of its likely link to the Iranian epidemic. Indeed, a search of all currently available GISAID sequences and metadata revealed no other complete genome sequences from patients with documented history of travel to or residence in Iran (as of 9 March 2020). A search of partial sequences identified two SARS-CoV-2 sequences which originated in Iran (413553/IRN/Tehran15AW/2020-02-28 and 413554/IRN/Tehran9BE/2020-02-23) spanning a 363 nt region of the viral nucleoprotein (N). Although short in length, these two sequences covered one of the informative SNPs defining this clade—T28688C, and both Iranian strains matched the sequences from patients with travel histories to Iran and grouped by phylogenetic analysis (Supplementary Figs. S1 and S2).

5. Discussion

Technological advancements and the widespread adoption of WGS in pathogen genomics have transformed public health and infectious disease outbreak responses (Popovich and Snitkin 2017). Previously, disease investigations often relied on the targeted sequencing of a small locus to identify genotypes and

Table 1. SARS-CoV-2 genomes sequenced in this study

GISAID ID	Virus name	Location	Collection date	Travel history
EPI_ISL_408976	408976/Australia/Sydney-2/2020-01-22	Sydney, Australia	22 January 2020	China
EPI_ISL_407893	407893/Australia/NSW01/2020-01-24	Sydney, Australia	24 January 2020	China
EPI_ISL_408977	408977/Australia/Sydney-3/2020-01-25	Sydney, Australia	25 January 2020	China
EPI_ISL_413490	413490/New_Zealand/01/2020-02-27	Auckland, New Zealand	27 February 2020	Iran
EPI_ISL_412975	412975/Australia/NSW05/2020-02-28	Sydney, Australia	28February 2020	Iran
EPI_ISL_413594	413594/Australia/NSW08/2020-02-28	Sydney, Australia	28 February 2020	SE Asia
EPI_ISL_413595	413595/Australia/NSW09/2020-02-28	Sydney, Australia	28 February 2020	SE Asia
EPI_ISL_413213	413213/Australia/NSW06/2020-02-29	Sydney, Australia	29 February 2020	Iran
EPI_ISL_413214	413214/Australia/NSW07/2020-02-29	Sydney, Australia	29 February 2020	None
EPI_ISL_413596	413596/Australia/NSW10/2020-02-28	Sydney, Australia	1 March 2020	SE Asia
EPI_ISL_413597	413597/AUS/NSW11/2020-03-02	Sydney, Australia	2 March 2020	Iran
EPI_ISL_413600	413600/AUS/NSW14/2020-03-03	Sydney, Australia	3 March 2020	None
EPI_ISL_413598	413598/AUS/NSW12/2020-03-04	Sydney, Australia	4 March 2020	Iran
EPI_ISL_413599	413599/AUS/NSW13/2020-03-04	Sydney, Australia	4 March 2020	Iran

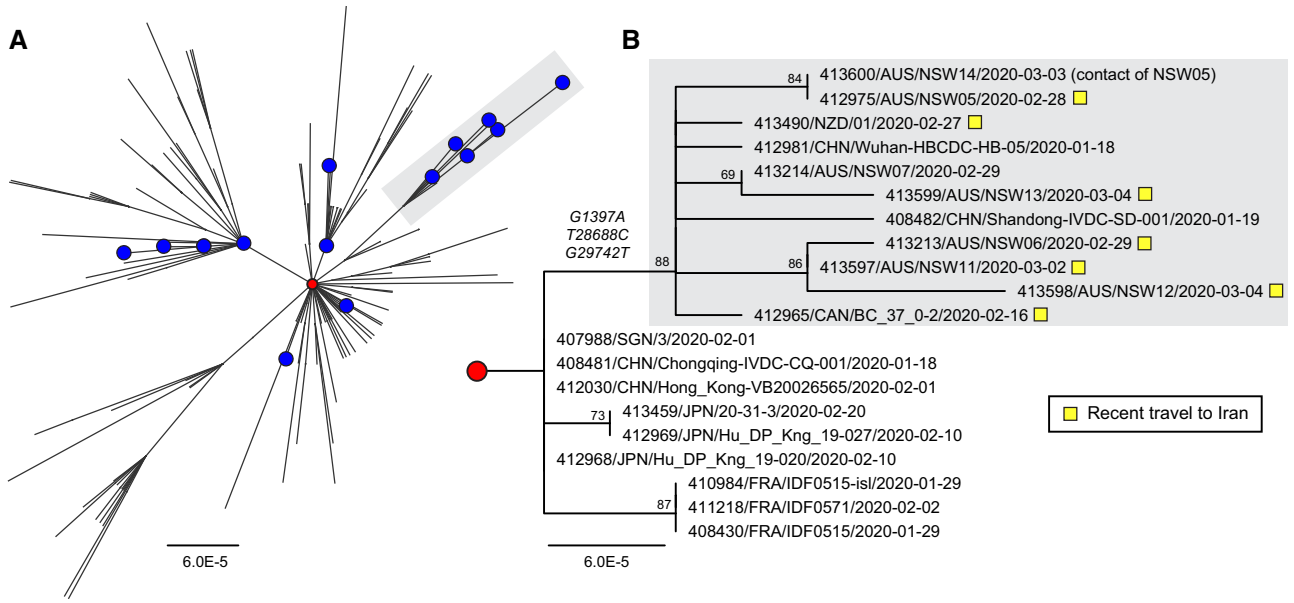

Figure 1. Phylogenetic analysis of SARS-CoV-2 genome sequences highlighting a clade of imported cases from Iran. (A) Global diversity of circulating SARS-CoV-2 strains including Australian sequences (blue circles, $n = 19$). The prototype strain Wuhan-Hu-1 is shown as a red circle. An emergent clade containing cases imported from Iran is highlighted with grey shading. (B) Sub-tree showing the informative branch containing imported Iranian cases (highlighted with yellow squares) and defined by substitutions at positions G1397A, T28688C, and G29742T. Node support is provided as bootstrap values of 100 replicates. For both (A) and (B), the scales are proportional to the number of substitutions per site.

infer patterns of spread along with epidemiological data (Dudas and Bedford 2019). As seen with the recent West African Ebola (Dudas et al. 2017) and Zika virus epidemics (Grubaugh et al. 2018), rapid WGS significantly increases resolution of diagnosis and surveillance thereby strengthening links between genomic, clinical, and epidemiological data (Grenfell 2004), and potentially uncovering outbreaks in unsampled locations (Grubaugh et al. 2019b). This advance improves our understanding of pathogen origins and spread that ultimately lead to stronger and more timely intervention and control measures (Grubaugh et al. 2019c). Following the first release of the SARS-CoV-2 genome (Wu et al. 2020), public health and research laboratories worldwide have rapidly shared sequences on public data repositories such as GISAID (Shu and McCauley 2017) ($n = 236$ genomes as of 9 March 2020) that have been used to provide near real-time snapshots of global diversity through public analytic and visualization tools (Hadfield et al. 2018).

Although all known cases linked to Iran are contained in this clade, it is important to note the presence of two Chinese strains sampled during mid-January 2020 from Hubei and Shandong provinces. It is expected that further Chinese strains will be identified that fall in this clade, and across the entire phylogenetic diversity of SARS-CoV-2 as this is where the outbreak started, including likely for the outbreak in Iran itself. However, while we cannot completely discount that the cases in Australia and New Zealand came from other sources including China, our phylogenetic analyses, as well as epidemiological (recent travel to Iran) and clinical data (date of symptom onset), provide evidence that this clade of SARS-CoV-2 is directly linked to the Iranian epidemic, from where genomic data are currently lacking. Importantly, the seemingly multiple importations of very closely related viruses from Iran into Australia suggest that this diversity reflects the early stages of SARS-CoV-2 transmission within Iran.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

The members of the nCoV-2019 Study Group also include Linda Donovan, Shanil Kumar, Tyna Tran, Danny Ko, Christine Ngo, Tharshini Sivaruban, Verlaine Timms, Connie Lam, Mailie Gall, Karen-Ann Gray, Rosemarie Sadsad, and Alicia Arnott. The authors acknowledge the Sydney Informatics Hub and the use of the University of Sydney's high-performance computing cluster, Artemis, and all the laboratories that have referred SARS-CoV-2 samples to the Centre for Infectious Diseases and Microbiology Laboratory Services, NSW Health Pathology – Institute of Clinical Pathology and Medical Research, Westmead Hospital. We would finally like to thank all the authors who have kindly shared genome data on GISAID, and we have included a table (Supplementary Table S2) outlining the authors and institutes involved. Data including the sequences in this study are available for download from https://www.gisaid.org/.

Author contributions

Study concept and design by J.S.E., E.C.H., and J.K. Sample processing and testing by I.C. and H.R. Sequencing and analysis by J.S.E., R.R., J.D.L., J.H., M.S., X.R., R.T., and E.C.H. Study coordination by K.B., J.W., R.B., N.G., M.V.O., V.S., S.C.C., S.M., T.C.S., D.E.D., and J.K. J.S.E. wrote the first manuscript draft with editing from E.C.H., J.D.L., R.R., T.C.S., V.S., and J.K. The final manuscript was approved by all authors.

Funding

This study was supported by the Prevention Research Support Programme funded by the New South Wales Ministry of Health and the NHMRC Centre of Research Excellence in Emerging Infectious Diseases (GNT1102962).

Conflict of interest: None declared.

References

Di Giallonardo, F. et al. (2018) 'Evolution of Human Respiratory Syncytial Virus (RSV) over Multiple Seasons in New South Wales, Australia', *Viruses*, 10: 476.

Dong, E., Du, H., and Gardner, L. (2020) 'An Interactive Web-Based Dashboard to Track COVID-19 in Real Time', *The Lancet Infectious Diseases*, 20: 30120–1.

Dudas, G., and Bedford, T. (2019) 'The Ability of Single Genes vs Full Genomes to Resolve Time and Space in Outbreak Analysis', *BMC Evolutionary Biology*, 19: 1–17.

—— et al. (2017) 'Virus Genomes Reveal Factors That Spread and Sustained the Ebola Epidemic', *Nature*, 544: 309–15.

Grenfell, B. T. (2004) 'Unifying the Epidemiological and Evolutionary Dynamics of Pathogens', *Science*, 303: 327–32.

Grubaugh, N. D. et al. (2018) 'Genomic Insights into Zika Virus Emergence and Spread', *Cell*, 172: 1160–2.

—— et al. (2019a) 'An Amplicon-Based Sequencing Framework for Accurately Measuring Intrahost Virus Diversity Using PrimalSeq and iVar', *Genome Biology*, 20: 8.

—— et al. (2019b) 'Travel Surveillance and Genomics Uncover a Hidden Zika Outbreak during the Waning Epidemic', *Cell*, 178: 1057–71.

—— et al. (2019c) 'Tracking Virus Outbreaks in the Twenty-First Century', *Nature Microbiology*, 4: 10–9.

Guindon, S., and Gascuel, O. (2003) 'A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood', *Systematic Biology*, 52: 696–704.

Hadfield, J. (2019) *ARTIC Network RAMPART* <https://github.com/artic-network/rampart> accessed 10 March 2020.

—— et al. (2018) 'Nextstrain: Real-Time Tracking of Pathogen Evolution', *Bioinformatics*, 34: 4121–3.

Katoh, K. (2002) 'MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform', *Nucleic Acids Research*, 30: 3059–66.

Loman, N., and Rambaut, A. (2020) *ARTIC Network Bioinformatics SOP* <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html> accessed 10 March 2020.

Lu, R. et al. (2020) 'Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding', *The Lancet*, 395: 565–74.

Popovich, K. J., and Snitkin, E. S. (2017) 'Whole Genome Sequencing-Implications for Infection Prevention and Outbreak Investigations', *Current Infectious Disease Reports*, 19: 15.

Quick, J. (2020) *nCoV-2019 Sequencing Protocol* <https://www.protocols.io/view/ncov-2019-sequencing-protocol-single-sample-bdbfi2jn> accessed 10 March 2020.

Shu, Y., and McCauley, J. (2017) 'GISAID: Global Initiative on Sharing All Influenza Data—From Vision to Reality', *Eurosurveillance*, 22: 30494.

World Health Organisation (2020a) Coronavirus Situation Report—8th *March 2020* <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200308-sitrep-48-covid-19.pdf? sfvrsn=16f7ccef_4> accessed 9 March 2020.

—— (2020b) *Coronavirus Disease (COVID-19) Technical Guidance: Laboratory Testing for 2019-nCoV in Humans* <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical guidance/laboratory-guidance> accessed 10 March 2020.

Wu, F. et al. (2020) 'A New Coronavirus Associated with Human Respiratory Disease in China', *Nature*, 579: 265–9.

Wuhan Municipal Health Commission. (2019) *Briefing on the Current Pneumonia Epidemic Situation in Our City 2019* <http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989> accessed 9 March 2020.