Phylodynamics reveals the role of human travel and contact tracing in controlling COVID-19 in four island nations

Jordan Douglas[1,2,†,*], Fábio K. Mendes[1,2,3,†,*], Remco Bouckaert[1,2], Dong Xie[1,2], Cinthy L. Jiménez-Silva[1,3], Christiaan Swanepoel[1,2], Joep de Ligt[4], Xiaoyun Ren[4], Matt Storey[4], James Hadfield[5], Colin R. Simpson[6], Jemma L. Geoghegan[4,7], David Welch[1,2] and Alexei J. Drummond[1,2,3]

[1]Centre for Computational Evolution, The University of Auckland, Auckland, New Zealand
[2]School of Computer Science, The University of Auckland, Auckland, New Zealand
[3]School of Biological Sciences, The University of Auckland, Auckland, New Zealand
[4]Institute of Environmental Science and Research, Wellington, New Zealand.
[5]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[6]School of Health, Victoria University of Wellington, Wellington, New Zealand
[7]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

[†]To whom correspondence should be addressed; Email: jordan.douglas@auckland.ac.nz, f.mendes@auckland.ac.nz.
[*]Authors contributed equally to this work.

MAIN TEXT

The work below has not been peer reviewed

Abstract

Most populated corners of the planet have been exposed to SARS-CoV-2, the coronavirus behind the COVID-19 pandemic. We examined the progression of COVID-19 in four island nations that fared well over the first three months of the pandemic: New Zealand, Australia, Iceland, and Taiwan. Using Bayesian phylodynamic methods, we estimated the effective reproduction number of COVID-19 in the four islands as 1–1.4 during early stages of the pandemic, and show that it declined below 1 as human movement was restricted. Our reconstruction of COVID-19's phylogenetic history indicated that this disease was introduced many times into each island, and that introductions slowed down markedly when the borders closed. Finally, we found that New Zealand clusters identified via standard health surveillance largely agreed with those defined by genomic data. Our findings can assist public health decisions in countries with circulating SARS-CoV-2, and support efforts to mitigate any second waves or future epidemics.

Introduction

Many continued hoping that the epidemic would soon die out and they and their families be spared. Thus they felt under no obligation to make any change in their habits, as yet. Plague was an unwelcome visitant, bound to take its leave one day as unexpectedly as it had come.

Albert Camus ("The Plague", 1947)

The respiratory tract illness caused by Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) known as Coronavirus Disease 2019 (COVID-19; Rothan and Byrareddy, 2020) has caused substantial global health and economic impact, and the global epidemic appears far from over. COVID-19 was first detected in the city of Wuhan, Hubei province (China) in late 2019; (Zhu et al., 2020). By 11 March 2020 there were 100,000 cases across 114 countries leading to the World Health Organization (WHO) declaring a pandemic (World Health Organization, 2020b). By the end of April, 3 million cases and 200,000 deaths had been confirmed (World Health Organization, 2020a), with over half of the world population under stay-at-home orders (Sandford, 2020).

COVID-19 has now been characterised extensively. The average number of people each case goes on to infect (known as the effective reproduction number, R_{e}), for example, has been estimated as 1–6 during the early stages of the outbreak (Liu et al., 2020; Binny et al., 2020; Alimohamadi et al., 2020). Because COVID-19 often causes mild symptoms (Day, 2020b,a), the difference between the true and reported number of cases could be of an order of magnitude (Li et al., 2020a; Lu et al., 2020). Fortunately, R_e has since declined in many locations to less than 1 as result of social distancing, travel restrictions, contact tracing, among other mitigation efforts (Binny et al., 2020; Chinazzi et al., 2020).

At the moment of writing, different parts of the world tell contrasting stories about their struggle against COVID-19. New Zealand, Iceland, and Taiwan, for example, have fared comparatively well with smaller infection peaks and lower excess mortalities compared to many mainland countries (World Health Organization, 2020a; FT Visual & Data Journalism team, 2020). Australia was close to elimination, but in late June a second uncontrolled outbreak began in the state of Victoria. Nevertheless the initial experience of these four countries was similar. They all responded to COVID-19 by closing their borders and requiring isolation or quarantine of those who were allowed through (Hale et al., 2020), and either restricting or discouraging human movement (Apple, 2020). These countries provide excellent case studies to characterise the effects of human travel on epidemic spread, a critical endeavor when facing the threat of a COVID-19 second wave and global endemicity.

Mathematical methods for studying epidemics are as diverse as the diseases they target, making varied

use of surveillance and genetic data, and ranging from models of mathematical epidemiology (Grassly and Fraser, 2008; Binny et al., 2020; Ferretti et al., 2020; Hethcote, 2000) to phylodynamic models (Volz et al., 2013). The latter enjoy the comparable timescales of epidemiological and evolutionary processes, which result from short viral generation times and high viral mutation rates (20–40 mutations per year for SARS-CoV-2; Lai et al., 2020; Li et al., 2020b; Rambaut, 2020). By modelling the transmission process as phylogenetic trees (Fig. 1), phylodynamic and molecular evolution models can be coupled to reconstruct the mutational history of viral genomes (Li et al., 2020b; Rambaut, 2020; Lai et al., 2020), elucidating the epidemiological processes that shaped them.

We have sequenced 217 SARS-CoV-2 genomes from New Zealand, and evaluated the effects of travel and movement restrictions on disease progression in New Zealand, Australia, Iceland, and Taiwan. Finally, we examined the spatiotemporal structure of New Zealand's COVID-19 epidemic, by comparing genetically identified clusters to those independently identified by public health surveillance.

Phylodynamic models

In the study of disease phylodynamics, a complete epidemic can be modelled as a rooted, binary phylogenetic trees whose N tips are all infected individuals (Fig. 1), and whose characteristics and drivers we wish to uncover. In a rapidly growing epidemic the age of this tree is a good proxy for the duration of the infection in the population. Because it is seldom possible to sample the entire infected population, one instead reconstructs a subtree $\mathcal{T}$ from a sample of cases of size n, where $n < N$. This sample provides n viral genomic sequences we refer to as D, as well as each sequence's sampling date, collectively represented by vector y. D and y thus constitute the input data, and are herein referred to as the "samples". In the case of structured phylodynamic models, samples in y are further annotated by "deme" – a geographical unit such as a city, a country, or a continent.

A Bayesian phylodynamic model is characterised by the following unnormalised posterior probability density function:

$$f(\mathcal{T}, \mu_c, \theta_s, \theta_\tau | D, y) \propto \mathrm{P}(D|\mathcal{T}, \mu_c, \theta_s)\, f(\mathcal{T}|\theta_\tau)\, f_{pr}(\mu_c, \theta_s, \theta_\tau), \tag{1}$$

where $\mathrm{P}(D|\mathcal{T}, \mu_c, \theta_s)$ is the phylogenetic likelihood of tree $\mathcal{T}$ given alignment D, whose sequences accumulate substitutions at rate μ_c under a substitution model with parameters θ_s. $f(\mathcal{T}|\theta_\tau)$ is the probability density function of the tree prior sampling distribution, given tree prior parameters θ_τ (in this study we

consider a range of phylodynamic models that differ in their tree prior; see below and Supplementary materials). Finally, $f_{pr}(\mu_c, \theta_s, \theta_\tau)$ is the prior distribution on the remaining parameters.

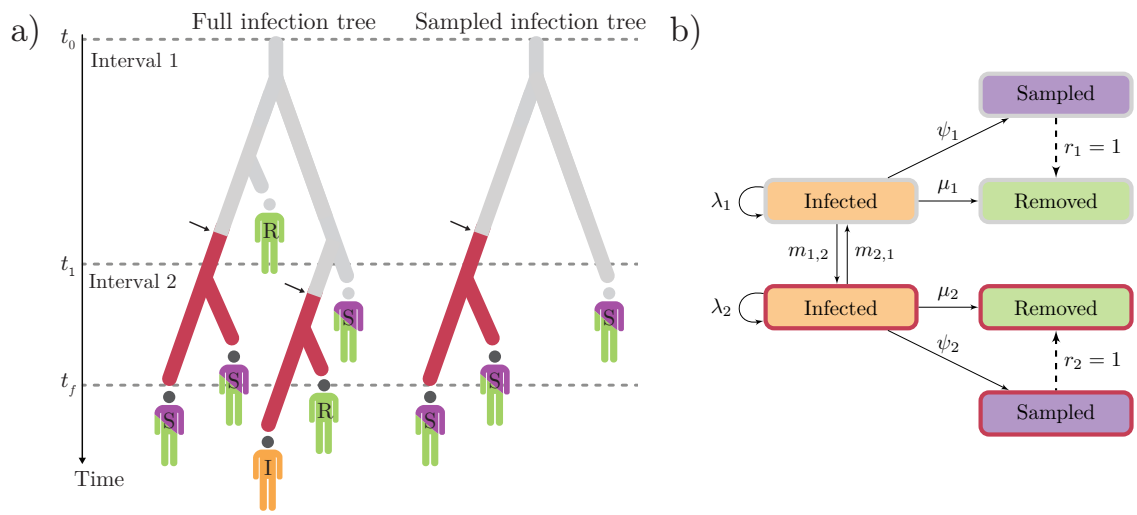

Fig. 1: Depiction of the MTBD model. (a) A full infection tree and a sampled infection tree ($\mathcal{T}$), with two time intervals (epochs) and two demes (indicated in gray and red). All lineages in the trees represent infected individuals (I). Branching corresponds to transmission, color change (arrows) to migration, and infected individuals may be removed (R), or sampled and then removed (S). (b) Transition diagram; solid and dashed arrows labelled with rates and probabilities, respectively.

Multitype birth-death model

The multitype birth-death (MTBD) model (Kühnert et al., 2016; Sciré et al., 2020) is a structured phylodynamic model that assumes infected ("**I**"; Fig. 1) hosts of viral lineages are organised into a set of demes ("types"), and migrate between them at rate m (Fig. 1). Infected individuals transmit the infection at birth rate λ, and can be sampled ("**S**"; Fig. 1) and sequenced at sampling rate ψ, after which they are removed ("**R**"; Fig. 1) with probability r. We set $r = 1$ by assuming (sampled) confirmed cases are isolated and are unlikely to cause further infections. Removal without sampling happens at death rate μ (Fig. 1). Here, "death" refers to anything causing hosts to no longer be infectious, including recovery, host death or behavioural changes (e.g., self-isolation).

The probability density function of $\mathcal{T}$ under the MTBD sampling distribution (Kühnert et al., 2016) and its parameters θ_τ is given by:

$$f(\mathcal{T}|\theta_\tau) = f(\mathcal{T}|\lambda, \mu, \psi, m, r, \pi_G, t, \mathcal{O}), \tag{2}$$

where π_G represents the frequency of lineages in each deme at the root, and $\mathcal{O}$ is tree $\mathcal{T}$'s age.

The MTBD model is reparameterised into a computationally and epidemiologically convenient form, with new parameters $R_\mathrm{e} = \frac{\lambda}{\mu+\psi r}$, $b = \mu + \psi r$ and $s = \frac{\psi}{\psi+\mu}$. R_e is the disease's effective reproductive number, which is closely related to the basic reproduction number (Stadler et al., 2013), with both being equal at the onset of the epidemic. b is the rate at which individuals become non-infectious, and s is the probability of an individual being a sample (out of all removed individuals). The phylodynamic tree prior under this parameterisation is:

$$f\big(\mathcal{T}|\theta_\tau\big) = f\big(\mathcal{T}|R_\mathrm{e}, b, s, m, r, \pi_G, t, \mathcal{O}\big). \tag{3}$$

Under our MTBD model, R_e, b, s, and m are estimated across a series of time intervals (and held constant within intervals) defined by time boundaries t, with $t_0 < t_1 < \ldots < t_f$. t is not estimated here (see below andS upplementary material for more details), but fixed at epidemiologically informed dates, or dates derived from a human mobility model; t_0 is the beginning of the infection, and t_f is both the last sampling time and the end of the last interval (Fig. 1). Time intervals are both parameter- and deme-specific.

Finally, we consider four independent geographical models, each characterised by two demes: a target island deme (New Zealand, Australia, Iceland, or Taiwan; referred to as the "$\mathcal{IS}$" deme) and a mutually exclusive "rest-of-the-world" deme ("$\mathcal{RW}$"), which contributes samples from other countries in the world. All $\mathcal{RW}$ parameters are considered nuisance parameters.

Alternative phylodynamic models

Phylodynamic analyses can be carried out with a range of different models. In order to determine how robust our estimates are to model choice, we analysed our SARS-CoV-2 genomic datasets (see below) with three models in addition to MTBD: (i) a simple discrete phylogeography model that treats demes in the same fashion as molecular sites ("DPG", Lemey et al., 2009), (ii) a discrete phylogeography model with two time intervals ("DPG2"), and (iii) a structured coalescent model ("SC", Müller et al., 2018). A brief description of these models and their results can be found in the Supplementary materials.

Results

SARS-CoV-2 genomic data and subsampling

We sequenced and assembled 217 New Zealand SARS-CoV-2 genomes that are now available on GISAID (Shu and McCauley, 2017), which provided the remaining 3000+ genome sequences analysed here. Due to computational limitations, we subsampled GISAID sequences as follows: (i) two different subsample sizes ("small" vs. "large"), with 200 and 500 $\mathcal{RW}$ sequences, and (ii) two sample selection methods ("time" vs. "active") that chose sequences uniformly across time, or proportionally to the number of active cases over time. Up to 250 sequences from the $\mathcal{IS}$ deme are included in each alignment (where more than 250 were available, we subsampled them using the respective sampling method).

Permutating the two options above yields four subsampling protocols (detailed in the Supplementary materials), which were applied to the four $\mathcal{IS}$ demes, generating a total of 16 SARS-CoV-2 alignments (30 kb long, 0.3% ambiguous sites on average). Each alignment consists of sequences from one of the four islands ($\mathcal{IS}$) and from the rest of the world ($\mathcal{RW}$).

Modelling human movement as tracked by mobile phone data

With the goal of characterising the decrease in human movement over the weeks our SARS-CoV-2 samples were collected, we employed a Bayesian hierarchical sigmoidal model for the analysis of mobile phone data (Apple, 2020; Fig. 2). This analysis revealed that human movement generally decreased in all four islands, with Taiwan being the first to show a declining trend among the islands, but also the least pronounced one. New Zealand, Australia, and Iceland underwent a marked decline in human movement during March 2020, with a slight recovery starting around mid-April.

This analysis informed the three epoch model for R_e. The first interval for R_e ranged from the origin t_0 to the first reported case in each $\mathcal{IS}$ at t_1. The second describes normal human mobility and ranged from t_1 to the date of mobility reduction t_2 (shown for each $\mathcal{IS}$ in Fig. 2). The third ranged from t_2 to the final sample at t_f and describes a period of reduced human movement.

SARS-CoV-2 phylodynamics

Inference under the MTBD and SC models suggests they are fairly robust to subsampling protocols, as opposed to DPG and DPG2, which yielded unrealistically small estimates of infection duration (i.e., tree age) and of viral introduction counts from the "active" method. Such model behavior is unsurprising and has been reported elsewhere (De Maio et al., 2015), although more realistic behavior from DPG and DPG2

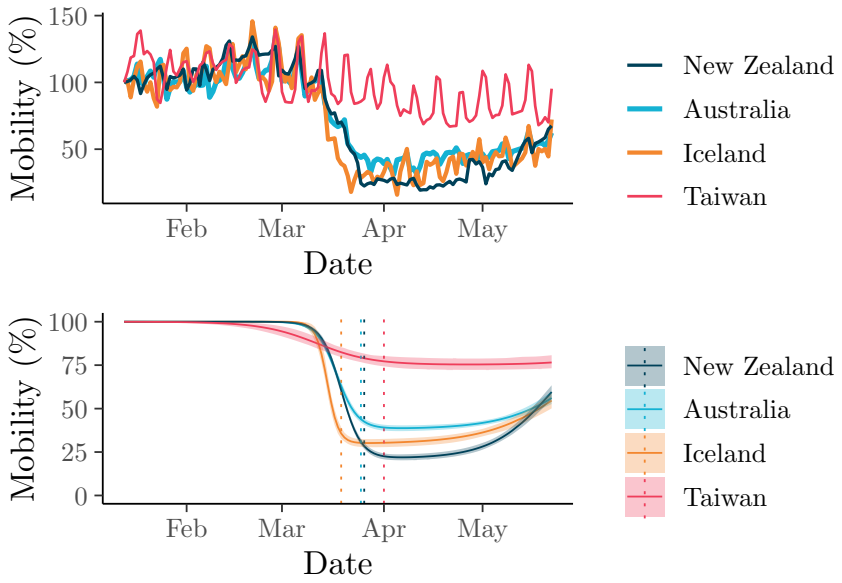

Fig. 2: Human walking movement according to cell phone mobility data (Apple, 2020) relative to day 1 (100%). Top: movement data from Apple. Bottom: fitted sigmoid model (solid lines show the median and shaded curves show the 2.5–97.5% quantile interval); vertical dotted lines correspond to the estimated mean date of mobility change (Supplementary materials).

was observed when "large" subsamples were used. Analyses under MTBD and SC were prohibitively time-demanding using the larger datasets. For consistency, we only present results from the "small-time" dataset here (but further results can be found in the Supplementary materials).

We estimated R_e as 1–2 during the early infection of each respective country and later at less than 1 after human mobility decreased (see credible intervals in Fig. 3). These estimates were not sensitive to sampling strategies. Each island $\mathcal{IS}$ saw multiple introductions of SARS-CoV-2 from the rest of the world $\mathcal{RW}$ (Fig. 4 and 5). Under MTBD, we estimated the mean number of sample sequence introductions to be 63, 87, 49, and 65 for New Zealand, Australia, Iceland, and Taiwan, respectively. Conversely, we estimated just one virus export event in each country. Similar results were observed using the SC model, with neither SC nor MTBD producing drastically different estimates between subsampling strategies.

We estimated the substitution rate of SARS-CoV-2 as approximately 9.4×10^{-4} substitutions per site per year, across the four phylodynamic models, the four islands, and the "small-time" subsampling method. These 16 (mean) estimates have a range of $(6.1 \times 10^{-4}, 12 \times 10^{-4})$, and are consistent with other estimates (Lai et al., 2020; Li et al., 2020b; Rambaut, 2020). By following the same pooling procedure, our estimates for the first transmission event range between 30 September and 25 December, 2019.

Comparison with contact tracing in New Zealand

The New Zealand Ministry of Health (NZMH) defines a significant COVID-19 cluster to comprise ten or more cases connected through transmission, and who are not all part of the same household (New Zealand Ministry of Health, 2020). Five of these clusters are represented by five or more of our samples (Fig. 4a and c), with a sixth non-significant cluster.

The most stringent definition of a phylogenetic cluster is a group of reciprocally monophyletic samples sharing the same outbreak identifying label. By this standard, only one of the six epidemiological clusters (i.e., six different labels), Matamata, was recovered from genetic data (Fig. 4).

Less stringently, one can define a phylogenetic cluster as a clade that contains all samples sharing a label, with only a single introduction into a focal deme being invoked. Under this definition (i) case exports are allowed, (ii) isolated labelled tips outside of the clade are considered type-1 contract tracing mismatches, and (iii) unlabelled tips that are both from the focal deme and placed within the clade are considered type-2 mismatches. Using this definition, all clusters identified by NZMH were recovered (among those represented by our samples). However, there were five type-1 mismatches: two from an Auckland-based cluster, one from Christchurch, one from Queenstown, and one from Bluff (shown by arrows in Fig. 4). We estimated 18 type-2 mismatches.

Discussion

Restricting human movement constrains COVID-19 spread

We explored the introduction and transmission history of COVID-19 in New Zealand, Australia, Iceland, and Taiwan – geographically isolated countries that were initially successful in preventing this disease from wreaking havoc as extensively it did in other parts of the world. Our results show that SARS-CoV-2 was introduced into each island on many independent occasions, corroborating the available epidemiological data from health officials (Supplementary materials). Viral imports rapidly declined after the respective islands closed their borders around 20 March 2020 (Fig. 4 and 5). Closing borders played a significant role in limiting the spread of COVID-19.

Mobile phone data revealed idiosyncratic human movement patterns (Fig. 2). During the early stages of the pandemic, human mobility rapidly declined in both New Zealand and Australia as a result of stay-at-home orders (Hale et al., 2020), as well as in Iceland despite no such official restraints. Taiwan, on the other hand, did not exhibit as much of a decline in human movement, nor did it enforce stay-at-home orders on the general population. A goal of this study was to characterise the extent to which this decline

in human movement could impact the spread of COVID-19.

We estimated effective reproduction numbers, R_e, as approximately 1–2 at the onset of the pandemic in all four islands. These estimates are not as high as in other countries where the pandemic was more widespread and where R_e was typically estimated in the range of 2-6 (Liu et al., 2020; Binny et al., 2020; Alimohamadi et al., 2020). Our low estimates reflect the fact that a large fraction of cases in these countries were travel-related imports, and that community transmission rates were low. R_e declined as human movement decreased in each of the countries studied, reaching values below 1.0 (Fig. 3).

The case of Taiwan illustrates that an extreme restriction of domestic human movement – officially mandated or otherwise – is not necessarily required for suppressing a pandemic like COVID-19. Instead, extensive contact tracing and the widespread use of face masks in Taiwan were likely key factors (Wang et al., 2020). However, a stringent response against COVID-19 greatly restricting mobility such as that adopted by New Zealand is likely to have contributed to the elimination of SARS-CoV-2 by May 9 (Cousins, 2020). New Zealand was the first country with over 1,000 confirmed cases to return to zero active cases, allowing it to re-open its economy, albeit with a still closed border. Most importantly, human morbidity and mortality were limited. Conversely, while Australia also saw a decline in human mobility and initially enjoyed some measure of success against COVID-19, its more lenient approach might underlie this country's ongoing struggle to fully eliminate SARS-CoV-2, as exemplified by the emergence of a second community-transmission led epidemic in the state of Victoria (Dexter, 2020).

Positive outcomes against this and similar pandemics will be dependent on a plethora of factors that must include human movement, but also population density, enforcement of social distancing, hygienic practices such as the use of face masks, the efficiency of contact tracing, high testing rates, and timely border closure. However, none of these should be seen as a silver bullet.

Phylodynamic modelling complements contact tracing

Identifying cases and the sources of their infections is an essential exercise in slowing infectious diseases such as COVID-19 (Hellewell et al., 2020), as recognised by the Director-General of the WHO in his pronouncement: "*You cannot fight a fire blindfolded. And we cannot stop this pandemic if we don't know who is infected*". Modern health surveillance strategies include medical databases containing travel history (Wang et al., 2020), the development of mobile phone applications (Ferretti et al., 2020), and the identification of outbreak subgroups known as "clusters", whose cases share the same origin (New Zealand Ministry of Health, 2020). At the core of health surveillance practice lies contact tracing, the goal of which is to pinpoint the origin of infections by reconstructing the travel history of confirmed cases, and identifying all

individuals with whom cases have come in contact.

Nonetheless, contact tracing has its limitations. Establishing the link between cases is restricted, for example, by the ability of cases to recall their recent contacts and travel history, which can lead to type-2 mismatches between epidemiological and genetic clusters. Furthermore, a case may have attended an event causally linked to a cluster but acquired their infection elsewhere and did not infect anyone at the event (type-1 mismatches). The implications of these types of error are manifold. If cluster sizes are underestimated, then so too is the rate of disease spread. If the extent of import-related cases is overestimated, then the impact of international air travel and the extent of community transmission cannot be fully accounted for. Indeed, our results suggest there are an additional 18 unclassified cases in New Zealand that could have been linked to known clusters but were not, and 5 cases that were linked to a cluster where they did not acquire or transmit the infection. However, by and large, our phylodynamic analysis is in agreement with conclusions reached by NZMH.

Overall, we have shown that contact tracing has been accurate among the cases considered, and succeeded to a large degree in identifying individuals belonging to the same infection cluster. We have demonstrated how the rapid real-time availability and assessment of viral genomic data can complement and augment health surveillance strategies. Genome sequencing is increasingly affordable and shared global surveillance resources such as NextStrain and GISAID are becoming more complete and useful. We advocate incorporating modern methods of genetic epidemiology into standard contact tracing and epidemic surveillance in order to ensure COVID-19 and future epidemics leave as promptly as they come.

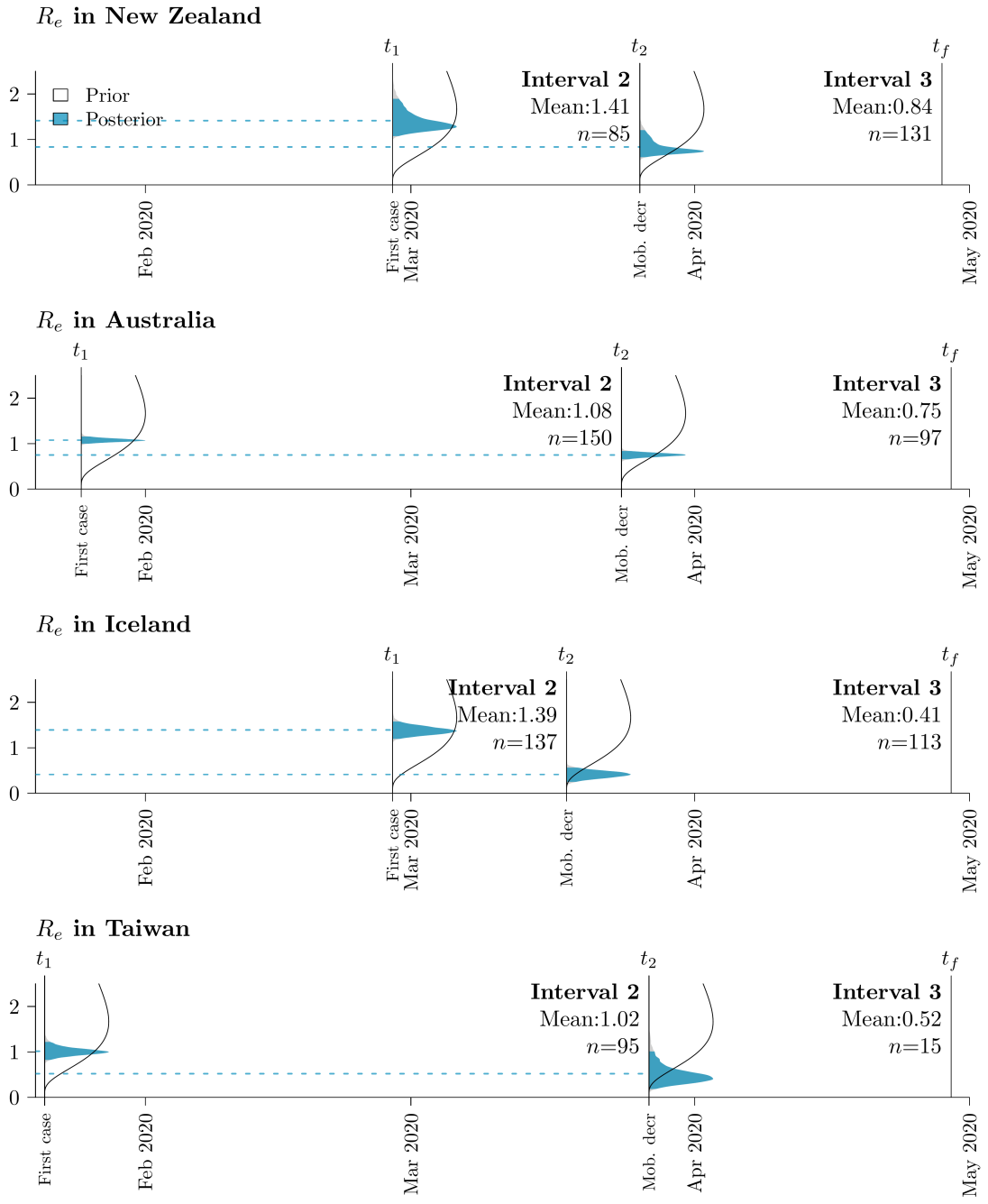

Fig. 3: R_e prior (white) and posterior (blue) probability distributions across islands and time intervals (n samples in each interval). Posterior means indicated by dashed line.

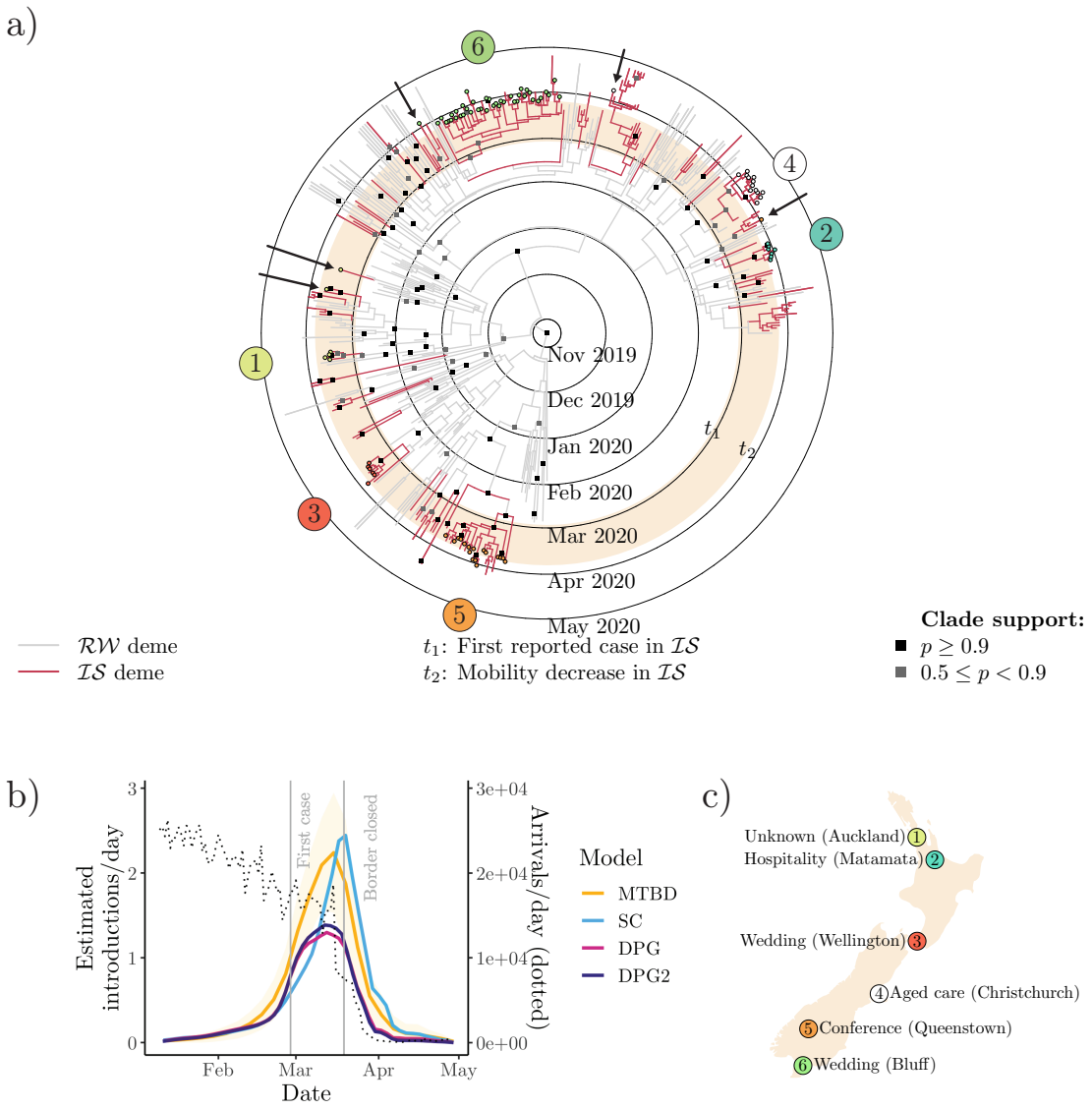

Fig. 4: (a) New Zealand COVID-19 maximum-clade-credibility tree (from "small-time" alignment). Samples are labelled according to outbreaks identified by New Zealand Ministry of Health (NZMH). Arrows indicate cases assigned to a cluster by NZMH which are not monophyletic with the rest of the cluster. (b) Number of SARS-CoV-2 introductions in New Zealand over time. (c) New Zealand cluster locations (with > 5 cases).

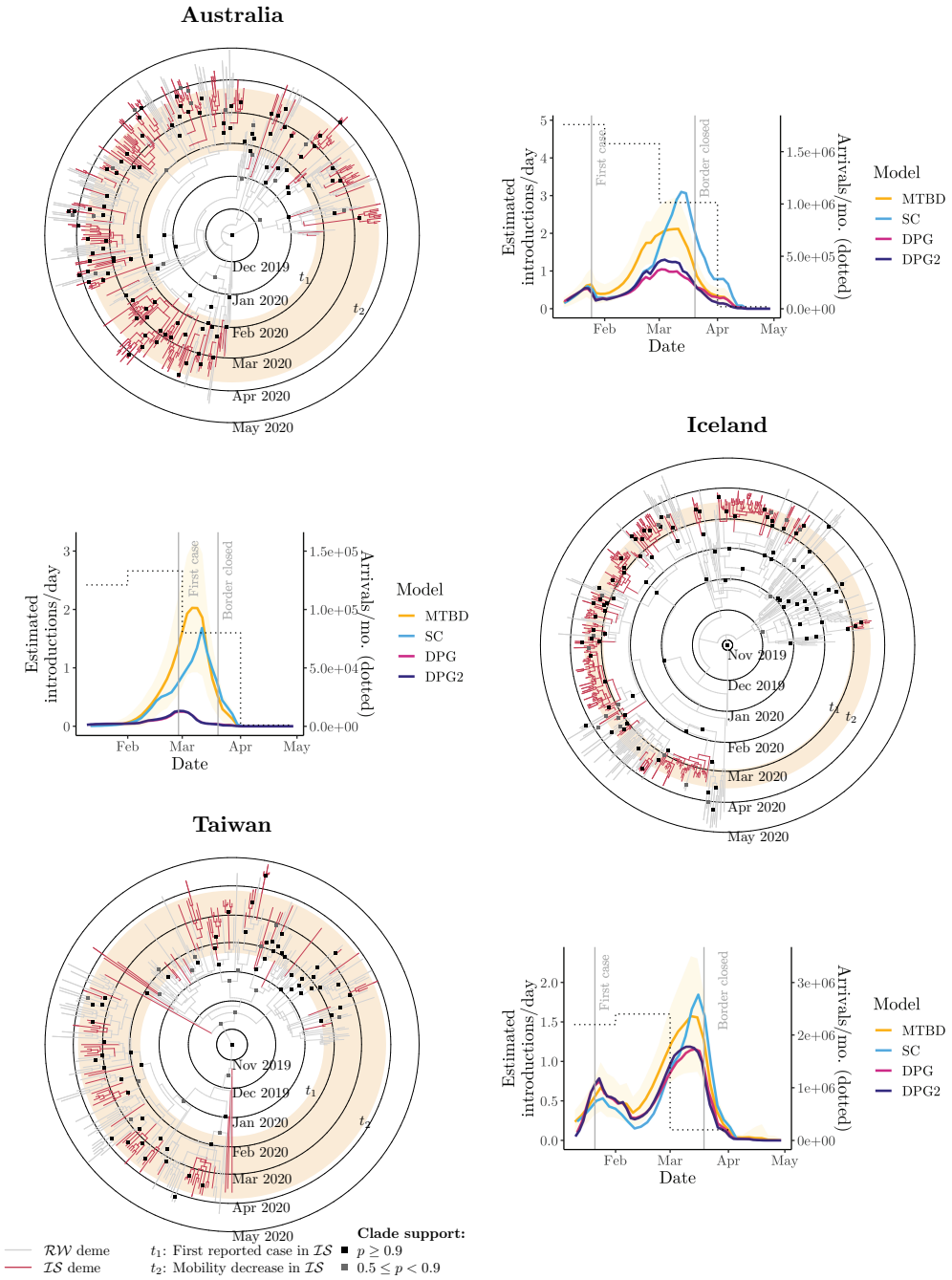

Fig. 5: Maximum-clade-credibility COVID-19 tree (from "small-time" alignment) for Australia (top), Iceland (middle), and Taiwan (bottom). Plots show the estimated number of SARS-CoV-2 introductions over time.

References

Alimohamadi, Y., M. Taghdir, and M. Sepandi. 2020. The estimate of the basic reproduction number for novel coronavirus disease (COVID-19): a systematic review and meta–analysis. J. Prev. Med. Public Health 53:151–157.

Apple. 2020. Mobility trends reports. [Online on `https://www.apple.com/covid19/mobility`; posted on 25 May 2020].

Bielejec, F., P. Lemey, G. Baele, A. Rambaut, and M. A. Suchard. 2014. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. Syst. Biol. 63:493–504.

Binny, R. N., S. C. Hendy, A. James, A. Lustig, M. J. Plank, and N. Steyn. 2020. Effect of alert level 4 on effective reproduction number: review of international COVID-19 cases. medRxiv .

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30:2114–2120.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, et al. 2019. BEAST 2.5: An advanced software platform for bayesian evolutionary analysis. PLoS Comput. Biol. 15:e1006650.

Bouckaert, R. R. and A. J. Drummond. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17:42.

Chinazzi, M., J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, et al. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368:395–400.

Cousins, S. 2020. New Zealand eliminates COVID-19. Lancet 395:P1474.

Day, M. 2020a. COVID-19: four fifths of cases are asymptomatic, China figures indicate. BMJ 369:m1375.

Day, M. 2020b. COVID-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. BMJ 368:m1165.

De Maio, N., C.-H. Wu, K. M. O'Reilly, and D. Wilson. 2015. New routes to phylogeography: a Bayesian structured coalescent approximation. PLoS Genet. 11:e1005421.

Dexter, R. 2020. Victoria records another 17 cases of coronavirus. The Age. [Online on https://www.theage.com.au/national/victoria/victoria-records-another-17-cases-of-coronavirus-20200623-p5558m.html; posted on 23 June 2020].

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22:1185–1192.

Eden, J.-S., R. Rockett, I. Carter, H. Rahman, J. De Ligt, J. Hadfield, M. Storey, X. Ren, R. Tulloch, K. Basile, et al. 2020. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. Virus Evol. 6:veaa027.

Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 368.

FT Visual & Data Journalism team. 2020. Coronavirus tracked: the latest figures as countries fight to contain the pandemic. Financial Times. [Online on https://www.ft.com/content/a26fbf7e-48f8-11ea-aeb3-955839e06441; posted on 28 May 2020].

Garrison, E. 2014. A C++ library for parsing and manipulating vcf files. [Online on https://github.com/vcflib/vcflib].

Gelman, A. et al. 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 1:515–534.

Grassly, N. C. and C. Fraser. 2008. Mathematical models of infectious disease transmission. Nat. Rev. Microbiol. 6:477–487.

Grubaugh, N. D., K. Gangavarapu, J. Quick, N. L. Matteson, J. G. De Jesus, B. J. Main, A. L. Tan, L. M. Paul, D. E. Brackney, S. Grewal, et al. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using primalseq and ivar. Genome Biol. 20:1–19.

Hadfield, J., C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, and R. A. Neher. 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34:4121–4123.

Hale, T., S. Webster, A. Petherick, T. Phillips, and B. Kira. 2020. Oxford COVID-19 government response tracker. Blavatnik School of Government. [Online on https://github.com/OxCGRT/covid-policy-tracker; posted on 4 June 2020].

Hellewell, J., S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, et al. 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health .

Hethcote, H. W. 2000. The mathematics of infectious diseases. SIAM Rev. 42:599–653.

Hoffman, M. D. and A. Gelman. 2014. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15:1593–1623.

Katoh, K., G. Asimenos, and H. Toh. 2009. Multiple alignment of DNA sequences with MAFFT. Pages 39–64 *in* Bioinformatics for DNA sequence analysis. Springer.

Kühnert, D., T. Stadler, T. G. Vaughan, and A. J. Drummond. 2016. Phylodynamics with migration: a computational framework to quantify population structure from genomic data. Mol. Biol. Evol. 33:2102–2116.

Lai, A., A. Bergna, C. Acciarri, M. Galli, and G. Zehender. 2020. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. J. Med. Virol. 92:675–679.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. PLoS Comp. Biol. 5.

Li, R., S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. 2020a. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 368:489–493.

Li, X., J. Zai, Q. Zhao, Q. Nie, Y. Li, B. T. Foley, and A. Chaillon. 2020b. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J. Med. Virol. 92:602–611.

Liu, Y., A. A. Gayle, A. Wilder-Smith, and J. Rocklöv. 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med. 27:1–4.

Loman, N., W. Rowe, and A. Rambaut. 2020. nCoV-2019 novel coronavirus bioinformatics protocol. [Online on `https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html`; posted on 23 January 2020].

Lu, F. S., A. Nguyen, N. Link, and M. Santillana. 2020. Estimating the prevalence of COVID-19 in the United States: three complementary approaches. medRxiv .

Müller, N. F. and R. Bouckaert. 2019. Coupled MCMC in BEAST 2. bioRxiv .

Müller, N. F., D. Rasmussen, and T. Stadler. 2018. MASCOT: parameter and state inference under the marginal structured coalescent approximation. Bioinformatics 34:3843–3848.

New Zealand Ministry of Health. 2020. COVID-19 – significant clusters. [Online on https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-current-situation/covid-19-current-cases/covid-19-significant-clusters; posted on 19 June 2020].

Rambaut, A. 2020. Phylodynamic analysis | 176 genomes | 6 mar 2020. [Online on http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356; posted on 6 March 2020].

Rothan, H. A. and S. N. Byrareddy. 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J. Autoimmun. Page 102433.

Sagulenko, P., V. Puller, and R. A. Neher. 2018. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 4.

Salvatier, J., T. V. Wiecki, and C. Fonnesbeck. 2016. Probabilistic programming in Python using PyMC3. PeerJ Comput. Sci. 2:e55.

Sandford, A. 2020. Coronavirus: half of humanity now on lockdown as 90 countries call for confinement. Euronews. [Online on https://www.euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou; posted on 3 April 2020].

Sciré, J., J. Barido–Sottani, D. Kühnert, T. G. Vaughan, and T. Stadler. 2020. Robust phylodynamic analysis of genetic sequencing data from structured populations. bioRxiv .

Shu, Y. and J. McCauley. 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. Euro. Surveill. 22:30494.

Stadler, T., D. Kühnert, S. Bonhoeffer, and A. J. Drummond. 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl. Acad. Sci. U.S.A 110:228–233.

Volz, E. M., K. Koelle, and T. Bedford. 2013. Viral phylodynamics. PLoS Comput. Biol. 9:e1002947.

Wang, C. J., C. Y. Ng, and R. H. Brook. 2020. Response to covid-19 in taiwan: big data analytics, new technology, and proactive testing. Jama 323:1341–1342.

World Health Organization. 2020a. Coronavirus disease (COVID-19) Situation Report – 101. [Online on https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200430-sitrep-101-covid-19.pdf?sfvrsn=2ba4e093_2; posted on 30 April 2020].

World Health Organization. 2020b. WHO director-general's opening remarks at the media briefing on COVID-19—11 march 2020. [Online on https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020; posted on 11 March 2020].

Zhu, N., D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, et al. 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382:727–733.