Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Estimating the age of the subfamily *Orthocoronavirinae* using host divergence times as calibration ages at two internal nodes

David T.S. Hayman *, Matthew A. Knox

Molecular Epidemiology and Public Health Laboratory, School of Veterinary Science, Massey University, New Zealand

**A R T I C L E   I N F O**

**Keywords:**
- Coronavirus
- Evolution
- Bats
- Biogeography

**A B S T R A C T**

Viruses of the subfamily *Orthocoronavirinae* can cause mild to severe disease in people, including COVID-19, MERS and SARS. Their most common natural hosts are bat and bird species, which are mostly split across four virus genera. Molecular clock analyses of orthocoronaviruses suggested the most recent common ancestor of these viruses might have emerged either around 10,000 years ago or, using models accounting for selection, many millions of years. Here, we reassess the evolutionary history of these viruses. We present time-aware phylogenetic analyses of a RNA-dependent RNA polymerase locus from 123 orthocoronaviruses isolated from birds and bats, including those in New Zealand, which were geographically isolated from other bats around 35 million years ago. We used this age, as well as the age of the avian-mammals split, to calibrate the molecular clocks, under the assumption that these ages are applicable to the analyzed viruses. We found that the time to the most recent ancestor common for all orthocoronaviruses is likely 150 or more million years, supporting clock analyses that account for selection.

1. Introduction

Orthocoronaviruses (family *Coronaviridae*, subfamily *Orthocoronavirinae*) are infectious agents of birds and mammals. In humans, they can cause mild illness and commonly colds, but emergent viruses can cause more severe disease, with Severe Acute Respiratory Syndrome (SARS) (Peiris et al., 2003), Middle Eastern Respiratory Syndrome (MERS) (Zaki et al., 2012) and Coronavirus disease 2019 (COVID-19) (Huang et al., 2020; Wang et al., 2020; Wu et al., 2020) all causing more severe disease and death in varying proportions of cases. The *Coronaviridae* were recently reclassified, specifically the subfamily *Coronavirinae* was renamed to *Orthocoronavirinae*, with many species formally recognized (Gorbalenya et al., 2020). Orthocoronaviruses are positive-sense RNA viruses and are classified into *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* genera (Gorbalenya et al., 2020; de Groot RJ et al., 2011). Alphacoronaviruses and betacoronaviruses are only found in mammals, whereas gammacoronaviruses and deltacoronaviruses mainly infect birds. SARS (Peiris et al., 2003), in particular, initiated the discovery of novel orthocoronaviruses in humans, domesticated animals, and wildlife (Poon et al., 2005; Snijder et al., 2003; Wevers and van der Hoek, 2009; Zaki et al., 2012). Bats and birds host the greatest diversity of these viruses and are

the likely natural ‘ancestral’ reservoirs of the viruses (Wertheim et al., 2013; Wong et al., 2019; Zhou et al., 2021). Previous studies have identified both evidence for possible orthocoronavirus – host codivergence and coevolution as well as recent cross-species transmission events (Leopardi et al., 2018; Zhang et al., 2020).

Molecular clock analyses of the RNA-dependent RNA polymerase (RdRp) gene and five other genomic regions using different models suggest a time of most recent common ancestor (tMRCA) for the orthocoronaviruses of either around 10,000 years ago (Woo et al., 2012) or 293 (95% CI, 190 to 489) million years ago (Wertheim et al., 2013). The large difference between these approaches was due to the first model using viral isolation (tip) dates and substitution models including the general time-reversible substitution model with a four-bin gamma rate distribution (GTR + Γ4), possibly in the absence of a temporal signal (Rieux and Balloux, 2016), and a second accounting for purifying selection. This was achieved through the development of models in HyPhy that model the substitutions using GTR + Γ4 and a branch site random effects likelihood (B8- REL) model (Pond et al., 2011) to account for variation in selection pressure across codon sites and phylogenetic lineages (Wertheim et al., 2013). These models reduce the effect of purifying selection that prevents the estimation of ages (Wertheim and Pond, 2011) through maintaining evidence of sequence homology after

* Corresponding author.
E-mail address: d.t.s.hayman@massey.ac.nz (D.T.S. Hayman).

https://doi.org/10.1016/j.virol.2021.08.004

Received 16 June 2021; Received in revised form 11 August 2021; Accepted 11 August 2021

Available online 14 August 2021

0042-6822/© 2021 Elsevier Inc. This article is made available under the Elsevier license (http://www.elsevier.com/open-access/userlicense/1.0/).
saturation at synonymous sites. For eukaryotic organisms, recent advances in phylogenetic analyses have allowed the use of fossils to calibrate these clocks (Gavrushkina et al., 2017; Heath, 2012), using fossil dates to calibrate nodes in the phylogenetic tree. Viruses are not fossilized and so tip calibration is usually used. However, endogenous viral elements and host species divergence ages have been used to estimate the age of other viruses, including single stranded, non-retroviral RNA viruses (Supplementary Table S1) (Belyi et al., 2016; Gifford et al., 2008; Gilbert and Feschotte, 2010; Han and Worobey, 2012; McGeoch and Cook, 1994; McGeoch et al., 1995; Suh et al., 2013, 2014; Taylor et al., 2010; Thézé et al., 2011). For example, the divergence of mammal hosts around 39–52 million years ago years ago with related endogenous filovirus elements lead to the estimation that filoviruses may be tens of millions years old, rather than the 10,000 years estimated by tip dates (Taylor et al., 2010).

Here we take advantage of biogeographic theory and mammalian speciation, including the unique features of phylogeography in New Zealand, to calibrate modeling approaches to estimate the age of bat and then representative bat and bird orthocoronaviruses. Specially, Alphacoronavirus RdRp RNA has been discovered in Mystacina tuberculata bats in New Zealand (Hall et al., 2014; Wang et al., 2015). New Zealand is estimated to have separated from other landmasses some 84 million years ago (Mortimer et al., 2017) and bats were, until the arrival of humans just 700 years ago between 1320 and 1350 (Walter et al., 2017), the only non-marine mammal present on the continent for nearly 35 million years. We assume there is an ancient, coevolutionary relationship between orthocoronaviruses and their bat or bird hosts for this analysis (Gorbalevna, 2008; Gorbalevna et al., 2006; Wertheim et al., 2013).

2. Materials and methods

2.1. Sequence data sets

Orthocoronavirus genomes (n = 123; Table 1) from all four genera were downloaded from GenBank in March 2020. Because only fragments of the RdRp gene (561 bp) were available from the New Zealand bats, we limited our analyses to this genomic region. This has additional benefits because this region is apparently free of recombinant segments (Wertheim et al., 2013) and is relatively conserved (Ziebuhr et al., 2001). However, we tested nucleotide sequences for evidence of recombination using DualBrothers (Minin et al., 2005). A total of 105 sequences from bat hosts were used for the initial analyses, and a further 18 sequences from bird hosts were included in subsequent tests (Table 1). All nucleotide sequences were aligned at the amino acid level using MAFFT version 7 employing the E-INS-i algorithm (Katoh and Standley, 2013).

2.2. Phylogenetic analyses

We chose to use amino acid sequences in our primary analyses but, due to the short length of the available sequences, we also re-ran analyses using the nucleotides. We used BEAST v1.10.4 (Suchard et al., 2018) and BEAST2 v2.6.3 (Bouckaert et al., 2019) to analyze the amino acid and nucleotide sequence alignments respectively. We did not use tip dates as the timescales mean all isolates are effectively contemporaneously sampled. We initially assumed a constant population size and a strict clock with an LG substitution model for amino acid sequences (only available in BEAST v1) and HKY substitution model as well as a four-bin gamma rate distribution (GTR + F4) with a proportion of invariant sites for nucleotides. After some initial tree topology checking to confirm that genera and the New Zealand and Australian bat-derived orthocoronaviruses were monophyletic, we put a calibration prior on the node that is the common ancestor of the New Zealand and Australian bat-derived orthocoronaviruses. To check the sensitivity of our assumptions we then also put a prior on the age of all the bird-derived

Table 1: Orthocoronavirus sequences analyzed in this study.

| Genus and published virus name | Host species | Accession | Year | Sampling country |
|-------------------------------|-------------|-----------|------|------------------|
| Alphacoronavirus               |             |           |      |                  |
| Miniopterus bat coronavirus 1  | Miniopterus spp. | AY864196 | 2004 | Hong Kong        |
| Bat coronavirus HKU7           | Miniopterus magnator | DQ249226 | 2005 | Hong Kong        |
| Scotophilus bat coronavirus 512 | Scotophilus kuhii | DQ648858 | 2005 | China            |
| Bat coronavirus HKU2           | Rhinolophus simus | EF203064 | 2006 | China            |
| Bat coronavirus 1B             | Miniopterus pusillus | EU20137 | 2006 | China            |
| Bat coronavirus 1A             | Miniopterus magnator | EU20138 | 2005 | China            |
| Bat coronavirus HKU8           | Miniopterus pusillus | EU20139 | 2005 | China            |
| Bat coronavirus                | Miniopterus schreibersii | GU190243 | 2008 | Bulgaria         |
| Bat coronavirus                | Miniopterus schreibersii | GU190244 | 2008 | Bulgaria         |
| Miniopterus bat coronavirus/Kenyaa/K33/2006 | Miniopterus inflatus | HQ728485 | 2006 | Kenya            |
| Coronavirus BtCoV/FP565/Art_jam/PAN/2010 | Artibeus jamaicensis | JQ731784 | 2010 | Panama           |
| Coronavirus BtCoV/KP356/Art_jam/PAN/2010 | Artibeus jamaicensis | JQ731786 | 2010 | Panama           |
| Bat coronavirus                | Miniopterus schreibersii | KF294269 | 2012 | China            |
| Bat coronavirus                | Miniopterus schreibersii | KF294270 | 2012 | China            |
| Bat coronavirus                | Miniopterus schreibersii | KF294271 | 2012 | China            |
| Bat coronavirus                | Miniopterus schreibersii | KF294275 | 2012 | China            |
| Mystacina coronavirus New Zealand/2013 | Mystacina tuberculata | KF515987 | 2013 | New Zealand      |
| Mystacina coronavirus New Zealand/2013 | Mystacina tuberculata | KF515988 | 2013 | New Zealand      |
| Mystacina coronavirus New Zealand/2013 | Mystacina tuberculata | KF515989 | 2013 | New Zealand      |
| Mystacina coronavirus New Zealand/2013 | Mystacina tuberculata | KF515990 | 2013 | New Zealand      |
| Alphacoronavirus BtCoV/MSTM2/Minnat/SA/2010 | Miniopterus cf. natalensis | KF843851 | 2010 | South Africa     |
| Alphacoronavirus BtCoV/VH_NC2/Neo.cap/SA/2012 | Neoacromia cf. capensis | KF843854 | 2010 | South Africa     |
| Alphacoronavirus BtCoV/GrNC1/Neo.cap/SA/2012 | Neoacromia cf. capensis | KF843855 | 2010 | South Africa     |
| Alphacoronavirus BtCoV/GrNC2/Neo.cap/SA/2012 | Neoacromia cf. capensis | KF843856 | 2010 | South Africa     |
| BtRF-Alphacov/Hub2013 | Rhinolophus ferrumequinum | KJ473807 | 2013 | China            |
| BtMs-Alphacov/GS2013 | Myotis sp. | KJ473810 | 2013 | China            |
| 229E-related bat coronavirus   | Hippipusserus abae | KT253270 | 2010 | Ghana            |
| 229E-related bat coronavirus   | Hippipusserus abae | KT253272 | 2010 | Ghana            |
| 229E-related bat coronavirus   | Hippipusserus cf. ruber | KT253273 | 2010 | Ghana            |
| 229E-related bat coronavirus   | Hippipusserus abae | KT253274 | 2010 | Ghana            |
| 229E-related bat coronavirus   | Hippipusserus cf. ruber | KT253297 | 2011 | Ghana            |
| 229E-related bat coronavirus   | Hippipusserus cf. ruber | KT253298 | 2011 | Ghana            |
| Miniopterus bat coronavirus     | Miniopterus schreibersii | KU343194 | 2012 | China            |

(continued on next page)
| Genus and published virus name | Host species | Accession | Year | Sampling country |
|-------------------------------|-------------|-----------|------|------------------|
| Bat coronavirus               | Miniopterus capensis | MG817491  | 2015 | South Africa     |
| Bat alphacoronavirus          | Neoromicia capensis | MG1817498 | 2015 | South Africa     |
| Alphacoronavirus sp.          | Neoromicia capensis | MG817499  | 2015 | South Africa     |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687934  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687937  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687938  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687941  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687942  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687943  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687944  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687945  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687946  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687948  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687954  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687955  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687956  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687957  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687958  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687960  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687961  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687961  | 2014 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687965  | 2015 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH687966  | 2015 | Vietnam          |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH921428  | 2016 | China           |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MH921429  | 2016 | China           |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MK211370  | 2017 | China           |
| Alphacoronavirus sp.          | Scotophillus kuhii   | MK211372  | 2017 | China           |
| Alphacoronavirus sp.          | Chalinolobus morio   | MN602059  | 2016 | Australia       |
| Alphacoronavirus sp.          | Chalinolobus morio   | MN602060  | 2016 | Australia       |
| Alphacoronavirus sp.          | Chalinolobus morio alphacoronavirus |          |      |                  |
| Alphacoronavirus sp.          | Miniopterus pusillus bat coronavirus | MN11518  | 2018 | China           |
| Alphacoronavirus sp.          | Scoto philus kuhii bat coronavirus 512-related | MN11521  | 2018 | China           |
| Beta-coronavirus              | Rhinolophus pearsoni | DQ071615  | 2004 | China           |
| Bat SARS coronavirus³         | Typhonectes pacificus | DQ648794  | 2005 | China           |
| Bat coronavirus/133/2005       | Pipistrellus abramus | EF065512  | 2006 | China           |
| Bat coronavirus HKU5          | Rousettus leschenaulti | HM211100 | 2006 | China           |
| Gamma-coronavirus             | Gallus gallus        | FJ904719  | 1991 | USA             |
| Infectious bronchitis virus   | Gallus gallus        | FJ904721  | 1972 | USA             |
| Turkey coronavirus            | Meleagris gallopavo  | GQ427173  | 2003 | USA             |
| Turkey coronavirus            | Meleagris gallopavo  | GQ427175  | 1994 | USA             |
| Turkey coronavirus            | Meleagris gallopavo  | GQ427176  | 1998 | USA             |
| Infectious bronchitis virus   | Gallus gallus        | GQ504724  | 1941 | USA             |
| Infectious bronchitis virus   | Gallus gallus        | GU393336  | 1954 | USA             |
| Turkey coronavirus            | Gallus gallus        | JF829660  | 2004 | China           |
| Turkey coronavirus            | Gallus gallus        | JF829660  | 2010 | China           |
| Duck coronavirus              | Pycnonotus jocosus   | FJ376619  | 2007 | China           |
| Infectious bronchitis virus   | Pycnonotus jocosus   | FJ376621  | 2007 | China           |

(continued on next page)
viruses and subsequently only the bird derived orthocoronaviruses, so in
total used two calibration points in three scenarios: a bat-only, a
bird-bat, and a bird-only. The bat-only calibration was tested on two
datasets, one complete (123 sequences from bat and bird hosts) and one
using only bat hosts (105 sequences). The ages we used were 35 million
years (2.5–97.5% quartiles = 30–42 MY) for the New Zealand bats,
based on the estimated divergence time of M. mystacinica from other bats
(Van den Bussche and Hoofer, 2000), and 150 million years (2.5–97.5%
quartiles = 134–167 MY) for the age of birds (Hu et al., 2009).
However, because New Zealand has one other living bat species, Chalinolobus
tuberculatus, we also test the sensitivity of this divergence time, and use
the age of C. tuberculatus, 17 million years (2.5–97.5% quartiles = 14–20 MY)
(Dool et al., 2016). We identify calibrations with the taxa (e.g. bird,
bat) and the age in superscript a, e.g. bat 17 MYA for the 35MYA calibration on the
New Zealand-Australian bat virus ancestor node. Lastly, we also used a
relaxed clock for each calibration scenario and general time-reversible
substitution model with a four-bin gamma rate distribution (GTR + I + F)
for each nucleotide scenario. Strict clock models were run for 10
million MCMC samples with a 10% burnin and sampling every 1000,
whereas the relaxed clocks require longer MCMC chains, so were run for
100 million and sampled every 10000. The analyses covered 186 amino
acid and 561 nucleotide sites. All xml files are available at https://github.com/dts-h2/coronavirus_ancestry and can be replicated. Logs were
visualized in Tracer 1.7.1 (Rambaut et al., 2018) and trees plotted in
FigTree 1.4.4 (Rambaut, 2012). We checked for overall host-virus coevolution using parafit (Legendre et al., 2002) in the ape R package
(Paradis and Schliep, 2019) using taxa (or sister taxa) from the TimeTree
database (Hedges et al., 2015) and amino acid model with the bat 35 and
bird 150 calibration (see Supplementary Fig. S1). Further manipulation
and visualization were performed in R v4.0.4 using beastie (du Plessis,
2020); ggplot2 (Wickham, 2016); ggmcmc (Fernandez-Marin, 2016);
stringr (Wickham, 2019); and ggdist (Burling, 2018) packages.

3. Results

3.1. Assumptions and model adequacy

We found no support for recombination across the RdRp fragment
used (see Supplementary Fig. S2). As anticipated, the limited length of
amino acid sequences did not provide information for the modelling of
some parameter and as a result several of the relaxed clock models
returned effective samples sizes (ESS) of <200. However, crucially the
ESS for tree heights (ages) was >1179 (range 1179–5036). Using
nucleotide sequences, all chains converged with ESS all >200, with all
ESS for tree heights (ages) > 479 (range 479–5775). Overall, the
nucleotide tree topology was well supported, with high posterior

---

Table 1 (continued)

| Genus and published virus name | Host species | Accession | Year | Sampling country |
|--------------------------------|--------------|-----------|------|------------------|
| Thrush coronavirus HKU12        | Turdus      | FJ76622   | 2007 | China            |
| Munia coronavirus HKU13         | Lonchura     | JQ065044  | 2007 | China            |
| White-eye coronavirus HKU16     | Zosterops    | JQ065040  | 2007 | China            |
| Sparrow coronavirus HKU17       | Passer montanus | JQ065045  | 2007 | China            |
| Magpie robin coronavirus HKU18  | Copyschus    | JQ065046  | 2007 | China            |
| Night-heron coronavirus HKU19   | Nycticorax   | JQ065047  | 2007 | China            |
| Wigeon coronavirus HKU20        | Anas Penelope | JQ065048  | 2007 | China            |
| Common-moorhen coronavirus HKU21| Gallinula   |          |      |                  |

a Species, strain/isolate names are in Supplemental Table S2.
support for many nodes. There was support for coevolution between viruses and hosts over the whole tree using parafit (p value = 0.03, Supplementary Fig. S3). The posterior distributions of the HKY and GTR were essentially the same and, as expected there is increasing uncertainty of tMRCA estimates and lower node support with deeper nodes (see Supplementary Table S3, Supplementary Figs. S4–S33). Herein only the amino acid sequences using the 35 MY bat prior (bat$^{35}$) are discussed, but all the results can be seen in Table 2 and the supplementary information.

3.2. Time to the most recent common ancestor

The strict clock analyses using the bat$^{17}$, bat$^{35}$, bat$^{17}$ and bird$^{150}$, bat$^{35}$ and bird$^{150}$ or bird$^{150}$ only calibration points and LG model, estimated bat orthocoronaviruses to be somewhere from 133 to 391 million years old (median values). These values overlapped with the estimates from all other models for the relaxed clocks (Fig. 1). The relaxed clock models had great uncertainty, for example for the bat$^{35}$ only calibration with both bat and bird viruses the estimate was 305MY (38–748 95% HPD). The strict clock estimates were 391MY (190–708 95% HPD). All other values are in Table 2. Overall, our calibrations led to the substitution rates ranging from means of 2.4 $\times 10^{-3}$ to 7.7 $\times 10^{-4}$ per MY (see Table 2, Supplementary Table S4 and Fig. S34).

For all analyses, the single younger bat calibration point (17 MY) led to less uncertainty and younger tMRCA estimates if used alone. For the strict clock analyses the differences between estimates when the bird calibration point was used led to non-overlapping 95% HPD. However, in all analyses the estimates for the tMRCA for bat orthocoronaviruses is older than bats themselves (around 50MY, Fig. 2) and always includes the estimated age of birds (150MY), the only exception being the most uncertain strict clock analysis mentioned above with only the 35MY NZ-Australian bat clade prior (bat$^{35}$), that estimates the viruses to be older. In all cases the estimated orthocoronavirus crown tMRCA is similar to those of bat orthocoronaviruses. The calibration points force the nodes to be monophyletic and in most analyses the maximum clade credibility (MCC) trees had bat (alpha- and betacoronaviruses) with one common ancestor and bird (gamma- and deltacoronaviruses) with another. However, with only the bat calibration point, both the relaxed and strict clocks placed the bird virus ancestors as ancestral to bat viruses, with the switch only once occurring when both calibration points were used using the relaxed clock and GTR+$\Gamma$ model. All 14 amino acid and 16 nucleotide trees with the 35 million year old prior and their node support, 95% HPD and virus names are provided in the supplementary information.

4. Discussion

Our results support previous findings that orthocoronaviruses evolved millions of years ago (Wertheim et al., 2013). Increasingly analyses suggest that many viruses are ancient. Amniotes (reptiles, birds, and mammals) are estimated from fossil and molecular evidence to be around 325 million years old (Blair and Hedges, 2005; Shedlock and Edwards, 2009), whereas birds share an ancestor around 150 million years and bats around 50 million years ago (Simmons et al., 2008;
We estimated tMRCA dates for orthocoronaviruses of somewhere from 133 to 391 million years ago (median estimates), older than bats and (mostly) birds ancestors, suggesting these viruses evolved prior to bat and possibly bird ancestors among earlier Amniotes. There is, of course, great uncertainty in our estimates (Table 2, Fig. 1), but our results were robust to changes in the use of calibration point position. While the values and uncertainty changed, the use of calibrations either closer to tree tips (bat17, bat35) or deeper in the tree (bird150) nearer the crown of all orthocoronaviruses all led to ages older than bats and, mostly, birds. The family Coronaviridae includes Letovirinae subfamily viruses from a frog (Bukhari et al., 2018) and metatranscriptomic sequencing has identified distinct and diverse Coronaviridae phylogenetic groups among jawless and bony fish, providing further evidence that these viruses have evolved with vertebrate hosts over millions of years (Miller et al., 2021; Mordecai et al., 2019).

Here we also assumed these viruses were exclusively bat and bird viruses. It is feasible that other hosts (for example porcine, rodent, etc.) will come to light since host switching does occur, as evidenced by the recent emergence of orthocoronaviruses causing SARS, MERS and COVID-19 in people and swine acute diarrhoea syndrome in pigs (Zhou et al., 2018), and the widespread distribution of other now endemic orthocoronaviruses, such as HCoV-229E. We excluded other viruses because of this, but our results may have influenced by such host switches among bat taxa. The monophyletic relationships of orthocoronaviruses in bats (i.e., alphacoronaviruses and betacoronaviruses) and birds (gammacoronaviruses and deltacoronaviruses), however, suggest these relationships are old and real (Chu et al., 2011; Wong et al., 2019; Woo et al., 2012). The sequencing of additional and/or more complete genomes of bat-derived orthocoronaviruses from islands, especially in the Pacific, may help support these findings.

It is most likely orthocoronaviruses arrived in New Zealand with the colonizing Mysticina (~35MYA) or Chalinolobus (~17MYA) bats, though the arrival through other non-bat species such as marine mammals or with humans is possible. Given the isolated population of Mysticina on a
New Zealand South Island in the Pacific Ocean the alphacoronavirus RdRp RNA was isolated from and the widespread distribution of bat alphacoronaviruses globally, this later arrival seems unlikely. Further, more recent host switching from non-bat (e.g. rodent) hosts is possible but highly unlikely given the close association of New Zealand bat alphacoronaviruses with Australian bat alphacoronaviruses and the very recent arrival of the first rodents to New Zealand. Rodents (specifically the Polynesian rat (Rattus exulans), known to Māori as kiore) only arrived in New Zealand from Polynesia with Māori, not via Australia, approximately 700 years ago, and other rodents only with Europeans. Therefore, it is more likely our results support those using the BS-REL tree calibration and that orthocoronaviruses have been infecting bats and birds for millions of years and, possibly, since their Anniote ancestors diverged approximately 300 million years ago.

There remain other issues to resolve for orthocoronaviruses. For example, while the partial RdRp gene we used was likely free of recent recombination, recombination among orthocoronaviruses is common and can even include genetic material from unrelated viruses, impacting any estimates relating to ancestry (Puskey et al., 2020). By using calibration points so far in the past, we automatically reduce the substitution rate (range 2.4 × 10⁻³ to 7.7 × 10⁻⁶ per MY). These rates had previously been estimated to be 10⁻⁵ to 10⁻⁶ mutations per site per replication (Eckerle et al., 2010) or around 10⁻³ substitutions per site per year (Hon et al., 2008; Wertheim et al., 2013). When clocks become unreliable, therefore, is unknown, though clearly analyses of more recent ancestries like SARS-coronaviruses are most reliable of all (Boni et al., 2020). Time dependent rates of molecular evolution have been observed over a wide range of taxa with short-term substitution rates exceeding long-term by an order of magnitude or more (Ho et al., 2011). Rates of molecular evolution in RNA viruses may span several orders of magnitude (Duffy et al., 2008) with this rapid accumulation of genetic differences in the short-term primarily due to the fidelity of the polymerase that is used during replication, though genome size and replication speed are also important factors. In the longer term, it is thought that strong purifying selection constrains the substitution rate (Duchêne et al., 2014; Holmes, 2003). Together, the evidence supports time varying rates of evolution and so substitution rates (including ours) need to be used and interpreted with caution (Duchêne et al., 2014).

In summary, our analyses using Bayesian evolutionary and tree analyses and mammalian tMRCA estimates have allowed us to make inferences about the age of orthocoronaviruses and support our intuition that orthocoronaviruses probably have an evolutionary history that matches their vertebrate hosts.

Funding
This work was supported by Royal Society Te Apārangi Rutherford Discovery Fellowship RDF-MAU1701

CRediT authorship contribution statement
David T.S. Hayman: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, preparation. Matthew A. Knox: Data curation, Formal analysis, Writing – review & editing.

Declaration of competing interest
☐ The authors declare that they have no known competing financial interests/personal relationships that could have appeared to influence the work reported in this paper.
☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Acknowledgements
Thanks to Prof Eddie Holmes, for his very helpful discussions and great generosity. We also thank two anonymous reviewers, and Virology editor, Prof Gorbatenya, who was exceptionally helpful.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2021.08.004.

References
Belyi, V.A., Levine, A.J., Skalka, A.M., 2010. Unexpected inheritance: multiple integrations of ancient Bornavirus and Ebola virus sequences in vertebrate genomes. PLoS Pathog. 6.
Blair, J.E., Hedges, S.B., 2005. Molecular phylogeny and divergence times of deuterostome animals. Mol. Biol. Evol. 22, 2275–2284.
Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.-Y., Ferry, B., Castoe, T., Rambaut, A., Robertson, D.L., 2020. Evolutionary origins of the SARS-CoV-2 coronavirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. 5, 1408–1417. https://doi.org/10.1038/s41564-020-0771-4, 11th ed.
Bouraquet, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourmest, M., Gavryushkina, A., Heled, J., Jones, G., Kuhner, D., De Maio, N., Matchencher, M., Mendes, F.K., Muller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.H., Xie, Z., Zhang, C., Stadler, T., Drummond, A.J., 2019. Beast 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15.
Bukhari, K., Mulley, G., Gulyaeva, A.A., Zhao, L., Shu, G., Jiang, J., Neuman, B.W., 2018. Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronaviridae, the proposed new Alphacentroviridae. Virology 524, 160–171.
Burling, J., 2018. Ggdistribute: A Ggplot2 Extension for Plotting Unimodal Distributions. R package version 1.0.1. https://github.com/iamamutt/ggdistribute/.
Chu, D.K., Leung, C.Y., Gilbert, M., Joyner, P.H., Ng, E.M., Tsaat, M.Y., Guan, Y., Peiris, J.S., Poon, L.L., 2011. Avian coronavirus in wild aquatic birds. J. Virol. 85, 12815–12820.
de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbatenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M., Talbot, P.J., Woo, P.C.Y., J.Z., 2011. Family Coronaviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, Y.R. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Academic Press, Ltd., London, United Kingdom, pp. 806–828.
Dool, S.E., O’Donnell, C.F.J., Monks, J.M., Puechmaille, S.J., Keith, G., 2016. Phylogeographic-based conservation implications for the New Zealand long-tailed bat, (Chalinolobus tuberculatus): identification of a single ESU and a candidate population for genetic rescue. Conserv. Genet. 17, 1067–1079.
du Plessis, L., 2020. Beastio: Functions for Pre- and Post-processing of BEAST and BEAST2 Files. R package version 0.2.5.
Duchêne, S., Holmes, E.C., Ho, S.Y., 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. In: Proceedings of the Royal Society B: Biological Sciences, vol. 281, p. 20140732.
Duffy, S., Shackleton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267–276.
Eckerle, L.D., Becker, M.M., Halpin, P.A., Li, K., Venter, E., Lu, X.T., Scherbakovs, S., Graham, R.L., Baric, R.S., Stockwell, T.B., Spiro, D.J., Denison, M.R., 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog. 6.
Fernandez-i-Marin, X., 2016. Gmcncalc: analysis of MCMC samples and bayesian inference. J Stat Softw 70, 7.
Gavryushkina, A., Heath, T.A., Kepka, D.T., Stadler, T., Welch, D., Drummond, A.J., 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Syst. Biol. 66, 57–73.
Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., Shafer, R.W., 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proc. Natl. Acad. Sci. Unit. States Am. 105, 20362–20367.
Gilbert, C., Feschtotte, C., 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8.
Gorbatenya, A.E., 2008. Genomic and evolutionary of the Nidovirales. In: Perlman, S., Gallagher, T., EJ, S. (Eds.), Nidovirales. ASM Press, Washington, DC, pp. 15–28.
Gorbatenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Laufer, C., Leontovich, A.M., Neuman, B.W., Penuaz, D., Perlman, S., Poon, L.L.M., Samborsky, D.Y., Sidoryk, L.A., Sola, S., Ziegler, J., Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5 (4), 536–544.
Gorbatenya, A.E., Enjuanes, L., Ziebauer, J., Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res. 117, 17–37.
Hall, R.J., Wang, J., Peacey, M., Moore, N.E., McInnes, K., Tompkins, D.M., 2014. New alphacoronavirus in Mystacinia tuberculata bats, New Zealand. Emerg. Infect. Dis. 20, 697–705.
Han, G.-Z., Worobey, M., 2012. An endogenous foamy-like viral element in the coelacanth genome. PLoS Pathog. 8.
Heath, T.A., 2012. A hierarchical bayesian model for calibrating estimates of species divergence times. Syst. Biol. 61, 793–809.
