On Classification and Taxonomy of Coronaviruses (Riboviria, Nidovirales, Coronaviridae) with special focus on severe acute respiratory syndrome-related coronavirus 2 (SARS-Cov-2)

Short title: On the Classification of Coronaviruses

Evgeny V. Mavrodiev [1,*], Melinda L. Tursky [2, 3], Nicholas E. Mavrodiev [4], Malte C. Ebach [5], David M. Williams [6].

[1] University of Florida, Florida Museum of Natural History, Museum Road and Newell Drive, Dickinson Hall, 301, Gainesville, FL, 32611, USA. evgeny@ufl.edu.

[2] Blood, Stem Cell, and Cancer Research Program; St Vincent's Centre for Applied Medical Research and Department of Hematology and BM Transplant, St Vincent's Hospital, Sydney; NSW, 2010; Australia. m.tursky@amr.org.au.

[3] St Vincent's Clinical School, Faculty of Medicine; UNSW Sydney; NSW, 2010; Australia.

[4] Cornerstone Academy, 1520 NW 34th St, Gainesville, FL 32605. nickmavrick11@gmail.com.

[5] Palaeontology, Geobiology and Earth Archives Research Centre (PANGEA), School of Biological, Earth and Environmental Sciences, UNSW, Kensington, NSW, 2052 Australia. mcebach@gmail.com.

[6] Department of Life Sciences, The Natural History Museum, London, SW7 5BD UK. DavidMyWilliams@gmail.com.

[*] Corresponding author.

Abstract: Coronaviruses are highly pathogenic and therefore important human and veterinary pathogens viruses worldwide *(1)*. Members of family Coronaviridae have previously been analysed phylogenetically, resulting in proposals of virus interrelationships *(2-5)*. However, available Coronavirus phylogenies remain unrooted, based on limited sampling, and normally depend on a single method (*2-11*). The main subjects of this study are the taxonomy and systematics of coronaviruses and our goal is to build the first natural classification of Coronaviridae using several methods of cladistic analyses *(12)*, Maximum Likelihood method, as well as rigorous taxonomic sampling, making the most accurate representation of Coronaviridae's relationships to date. Nomenclature recommendations to help effectively incorporate principles of binary nomenclature into Coronaviridae taxonomy are provided. We have stressed that no member of *Sarbecovirus* clade is an ancestor of SARS-Cov-2, and humans are the only known host.

Keywords: Coronaviridae, cladistics, rooted phylogenetic trees, taxonomy, binary nomenclature, generic circumscription, SARS-Cov-2.

One Sentence Summary: Multiple comprehensive phylogenetic analyses of all coronavirus species enabled testing of critical proposals on virus interrelationships.

Introduction

Coronaviridae is a family of the order Nidovirales (realm Riboviria) (*13, 14*). Accordingto the current summaries of International Committee on Taxonomy of Viruses (ICTV) (*13-16*) this family is divided into two subfamilies - Letovirinae and Orthocoronavirinae. These two subfamilies circumscribe five genera: monotypic genus *Alphaletovirus* of the subfamily Letovirinae (with single species *Microhyla letovirus* 1), and four non-monotypic genera of subfamily Orthocoronavirinae, namely:

1. genus *Alphacoronavirus* with 12 subgenera: *Colacovirus* (monotypic), *Decacovirus* (two species), *Duvinacovirus* (monotypic), *Luchacovirus* (monotypic), *Minacovirus* (two species), *Minunacovirus* (two species), *Myotacovirus* (monotypic), *Nyctacovirus* (monotypic), *Pedacovirus* (two species), *Rhinacovirus* (monotypic), *Setracovirus* (two species) and *Tegacovirus* (monotypic) (Table S1);
2. genus *Betacoronavirus* with five subgenera: *Embecovirus* (four species), *Hibecovirus* (monotypic), *Merbecovirus* (four species), *Nobecovirus* (two species) and *Sarbecovirus* (the number of species is under review) (Table S1);
3. genus *Gammacoronavirus* with two monotypic subgenera: *Cegacovirus* and *Igacovirus* (Table S1);
4. genus *Deltacoronavirus* with three monotypic subgenera *Andecovirus*, *Herdecovirus* and *Moordecovirus* and one non-monotypic subgenus *Buldecovirus* (four species) (Table S1).

The viruses of Coronaviridae, such as the severe acute respiratory syndrome (SARS) and related Middle East respiratory syndrome (MERS), are highly pathogenic (*1*). The recently

discovered virus SARS-Cov-2, which is also a member of this family, causes the Coronavirus disease 2019 (COVID-19) that was declared a pandemic by WHO in March 2020. Eight months later the total cases of COVID-19 are approaching 40,100,000 and cumulative deaths have exceeded 1 127 000 (www.who.int, October 2020). Thus, Coronaviridae is of high medical importance worldwide. Such importance has resulted in an urgency to further understand the relationships within the coronavirus family, and the viruses most closely related to SARS-Cov-2. However, phylogenetic analyses of Coronaviridae to date have remained in question due to the use of limited or arbitrary taxonomic sampling, and/or the use of unrooted trees (for example (*2, 3, 6, 8-11*), among others). These issues can result in bias and lack vital hierarchical information needed to understand the critical relationships between the viruses within this family. To address this urgent need, this study has undertaken a comprehensive taxonomic analysis of the family Coronaviridae, using the best taxonomic sampling that is based on all ICTV-approved genomes of all known coronaviruses. Additional viruses were also included to enable testing of the veracity of recently published proposals, included recently identified viruses such as SARS-Cov-2.

Almost all available phylogenetic studies of coronaviruses have been based on the parametric Maximum Likelihood (ML) method. Thus, in order to increase the accuracy of our current analyses and resolve the relationships within Coronaviridae, we have used ML method as well as non-parametric cladistic approaches:

1. standard maximum parsimony (MP),
2. three-taxon statement analysis (3TA), and the
3. Average Consensus (AC) analysis as applied to the array of the maximal relationships.

The last two methods have not previously been used to resolve relationships within the Coronaviridae.

The produced trees (Figure 1, Figures S1 - S5) enabled testing for monophyly of all available non-monotypic genera and infrageneric taxa of Coronaviridae to determine whether they

are representative of a natural hierarchy. If so, this validates the estimation of the relationships between all of the available genera/infrageneric taxa or particular virus species of Coronaviridae (incl. SARS-Cov-2).

The first important proposal assessed herein is that the newly described monotypic genus *Alphaletovirus* (Table S1), a member of subfamily Letovirinae, is a sister group of family Coronaviridae (*2*).

The second proposal assessed was by Tokarz *et al*. (*4*) who suggested that the general relationship within Coronaviridae subfamily Orthocoronavirinae is a simple hierarchy: ((({*Alphacoronavirus*},{*Betacoronavirus*}){*Gammacoronavirus*}){*Deltacoronavirus*}). Unfortunately Tokarz *et al*. (*4*) based their suggestion on limited taxonomic sampling. We would name this proposition the (((A, B) Γ) Δ) hypothesis of the general relationship within the subfamily.

Thirdly, the subgenus *Hibecovirus* has been placed as sister taxon of the presumably monotypic subgenus *Sarbecovirus* (for instance (*3, 5*)). This proposal is again based on unrooted phylogenies, so it has been tested herein with special attention to the placement of SARS-Cov-2.

In the fourth proposal, the close relationship of SARS-Cov-2 with bat coronaviruses RaTG13 *(21, 25)* and RmYN02 *(24, 25)* as well as with pangolin coronavirus (reviewed in (*17*)) is also tested.

In addition, within the context of our findings, some critical comments are made to the nomenclature of the family as well as to the general patterns of hosting within Coronaviridae, including newly discovered coronavirus SARS-Cov-2.

Results

All four trees have several major consistencies, as shown in the Strict Consensus (Figure 1 and Figure S5, see also Figures S1 – S4).

a. General pattern of the relationships within Coronaviridae and the placement of *Microhyla letovirus* 1 (Figure 1, Figure S1 – S5, Table S1).

The results of all analyzes have demonstrated that the hierarchy ((({*Alphacoronavirus*}, {*Betacoronavirus*}){*Gammacoronavirus*}){*Deltacoronavirus*}), with *Microhyla letovirus* 1 (subgenus *Milecovirus*, genus *Alphaletovirus*, subfamily Letovirinae) which has been defined as its sister taxon, form a general pattern of the relationship within Coronaviridae.

b. {*Alphacoronavirus*} clade (Figure 1, Figure S1 – S5, Table S1).

{*Alphacoronavirus*} clade includes all known species of the subgenus *Alphacoronavirus.* Within this clade, all four trees showed that the following taxa are sisters:

a. *Minacovirus* 1 and 2 (Ferret coronavirus and Mink coronavirus 1);

b. *Pedacovirus* 1 and 2 (Porcine epidemic diarrhea virus and *Scotophilus* bat coronavirus 512);

c. *Setracovirus* 1 and 2 (Human coronavirus NL63 and NL63-related bat coronavirus BtKYNL63-9b);

d. *Decacovirus* 1 and 2 (Bat coronavirus HKU10 and *Rhinolophus ferrumequinum* alphacoronavirus HuB-2013);

e. *Minunacovirus* 1 and 2 (Miniopterus bat coronavirus 1 and Miniopterus bat coronavirus HKU8).

Thus, within {*Alphacoronavirus*} clade, we were able to find five smaller clades (subclades): {*Decacovirus*}, {*Minacovirus*}, {*Minunacovirus*}, {*Pedacovirus*} and

{*Setracovirus*}with the exact correspondence of each of these clades to the previously established subgenera of the genus *Alphacoronavirus.*

From our results (Figure 1, Figure S1 – S5, Table S1) it is also clear, that

a. monotypic subgenus *Colacovirus* (Bat coronavirus CDPHE15) is a sister of *Pedacovirus* clade in all of the trees;

b. monotypic subgenus *Duvinacovirus* (Human coronavirus 229E) is a sister of the *Setracovirus* clade in all of the trees, and

c. monotypic subgenus *Tegacovirus* (Alphacoronavirus 1) is a sister of the *Minacovirus* clade in all of the trees.

The phylogenetic placement of monotypic subgenera *Luchacovirus* (Lucheng Rn rat coronavirus), *Myotacovirus* (*Myotis ricketti* alphacoronavirus Sax-2011) and *Rhinacovirus* (*Rhinolophus* bat coronavirus HKU2) depends on the method of the analyses (Figure 1, Figure S1 – S5, Table S1). It is worth stressing, however, that all of the cladistic methods (Figure S1 – S3), but not the ML method (Figure S4) have placed *Luchacovirus* as a sister of the {*Alphacoronavirus*}.

c. {*Betacoronavirus*} clade (Figure 1, Figure S1 – S5, Table S1).

{*Betacoronavirus*} clade includes all known species of the subgenus *Betacoronavirus*.

All of the analyses argue in favor of the general simple hierarchical relationship ({*Embecovirus*} ({*Merbecovirus*} ({*Nobecovirus*} (*Hibecovirus*, {*Sarbecovirus*})))) within *Betacoronavirus* clade (Figure 1, Figures S1 - S5).

The relationships of four species of the subgenus *Embecovirus* (1-4) (Betacoronavirus 1, China Rattus coronavirus HKU24, Human coronavirus HKU1 and Murine coronavirus), that

formed a clade with the same name {*Embecovirus*}, the sister of {*Betacoronavirus*} clade, depend on the method of the analysis (Figure 1, Figures S1 - S5; Table 1).

All four species of subgenus *Merbecovirus* form a clade {*Merbecovirus*}. In all trees, the *Merbecovirus* 1 (Hedgehog coronavirus 1) is a sister to the clade that contains three other members of the subgenus *Merbecovirus*: namely, *Merbecovirus* 2 (Middle East respiratory syndrome-related coronavirus (MERS)), *Merbecovirus* 3 (*Pipistrellus* bat coronavirus HKU5), and the constant sister of the later *Merbecovirus* 4 (*Tylonycteris* bat coronavirus HKU4) (Figure 1, Figures S1 - S5; Table S1).

The {*Sarbecovirus*} clade that corresponds to the subgenus *Sarbecovirus* includes the viruses of severe acute respiratory syndrome-related coronavirus (SARS), the newly discovered monophyletic SARS-Cov-2 (two accessions have been included to the analyses, SARS-Cov-2a and SARS-Cov-2b), as well as CoV-ZC45 and SARS Cov ZS B.

Clade {*Sarbecovirus*-SARS plus SARS Cov ZS B} (SARS-Clade of the Figure 1) is a sister of the remaining species of {*Sarbecovirus*}.

Depending on the analysis, either bat coronaviruses RmYN02 or RaTG13 have been placed as a sister of SARS-Cov-2. The pangolin coronavirus (isolate MP789) has been defined as a sister of the clade {RmYN02 plus RaTG13 plus SARS-Cov-2} in all of the analyses.

All trees define the monotypic subgenus *Hibecovirus* (Bat Hp-betacoronavirus) as a sister of {*Sarbecovirus*} clade.

Two members of subgenus *Nobecovirus*, namely *Nobecovirus* 1 (*Rousettus* bat coronavirus GCCDC1) and *Nobecovirus* 1 (*Rousettus* bat coronavirus HKU9) are sister taxa (Figure 1, Figures S1 - S5; Table S1).

d. {*Gammacoronavirus*} clade (Figure 1, Figure S1 – S5, Table S1).

Two subgenera of the genus *Gammacoronavirus*, namely subgenus *Cegacovirus* (with the single species Beluga whale coronavirus SW1) and subgenus *Igacovirus* (with a single species Avian coronavirus) formed a clade {*Gammacoronavirus*} in all of the analyses (Figure 1, Figures S1 - S5).

e. {*Deltacoronavis*} clade (Figure 1, Figure S1 – S5, Table S1).

Five subgenera of genus *Deltacoronavis*, namely subgenus *Andecovirus* (with single species Wigeon coronavirus HKU20), subgenus *Buldecovirus* (1-4) (with four species: Bulbul coronavirus HKU11, Coronavirus HKU15, Munia coronavirus HKU13 and White-eye coronavirus HKU16), subgenus *Herdecovirus* (with one species Night heron coronavirus HKU19) and subgenus *Moordecovirus* (with single species Common moorhen coronavirus HKU21), have formed the {*Deltacoronavis*} clade in all of the analyzes. Also, all of the analyses argue in favor of the simplest hierarchy of the relationships within this clade: (*Andecovirus* (*Herdecovirus* (*Moordecovirus* ({*Buldecovirus*})))). Within {*Buldecovirus*} clade *Buldecovirus* 2 (Coronavirus HKU15) and *Buldecovirus* 3 (Munia coronavirus HKU13) appeared to be sisters in all of the analyses, the relationships between the other members of the {*Buldecovirus*} clade depends on the method of the analysis. Monotypic subgenus *Andecovirus* is a sister of {*Deltacoronavis*} clade.

f. On monophyly of non-monotypic taxa of Coronaviridae (Figure 1, Figure S1 – S5, Table S1).

As is perhaps clear from above, all four trees show that all four genera of subfamily Orthocoronavirinae (family Coronaviridae), namely *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus*, and *Gammacoronavirus*, are monophyletic (Figure 1, Figures S1 - S5).

All non-monotypic subgenera of all four genera of subfamily Orthocoronavirinae (family Coronaviridae), namely subgenus *Decacovirus* (genus *Alphacoronavirus*), subgenus *Minacovirus* (genus *Alphacoronavirus*), subgenus *Minunacovirus* (genus *Alphacoronavirus*), subgenus *Pedacovirus* (genus *Alphacoronavirus*), subgenus *Setracovirus* (genus *Alphacoronavirus*), subgenus *Embecovirus* (genus *Betacoronavirus*), subgenus *Merbecovirus* (genus *Betacoronavirus*), subgenus *Nobecovirus* (genus *Betacoronavirus*), and subgenus *Buldecovirus* (genus *Deltacoronavirus*) are monophyletic in all of the analyses (Figure 1, Figures S1 – S5).

Discussion

Several phylogenies of Coronaviridae (or parts thereof) have been published; however almost all of these remain unrooted (for example (*2, 3, 6, 8-11*) among others), or if the phylogenetic tree occasionally appears as rooted (for instance (*4, 7*)) the taxonomic sampling of such studies remain incomplete or even arbitrary. We would like to stress that this study seeks to produce comprehensive rooted phylogenetic Coronaviridae trees and that the analyses herein have used the best taxonomic summaries of family Coronaviridae currently established by ICTV (2019) (*1-4*).

a. Rigorous rooted trees validate tested proposals of the relationships within Coronaviridae

Rooted trees produced by the four methods herein allow for the effective testing of various proposals regarding the relationships of viruses within the Coronaviridae family.

1. Our results argue in favor of Bukhari *et al*. (*2*) proposal regarding the new family Abyssoviridae of the order Nidovirales (current monotypic subfamily Letovirinae with subgenus *Milecovirus*) (Table S1), which was unfortunately based on an unrooted phylogenetic tree and limited taxonomic sampling. After Bukhari *et al*. (*2*) we also found that two subfamilies of Coronaviridae, namely Letovirinae (family Abyssoviridae in Bukhari *et al*. (*2*)) and Orthocoronavirinae are sisters (Figure 1, Figures S1 - S5). This solution is consistent with the familial rank of both taxa (*2*). We would recommend to accept the monotypic subfamily Letovirinae at the familiar rank (*2*).
2. Tokarz *et al*. (*4*) proposed the (((A, B) Γ) Δ) hypothesis of the general pattern of relationship within subfamily Orthocoronavirinae; however this was again based on limited taxonomic sampling and an unrooted network. Our results clearly argue in favor of this hypothesis. Keeping in mind the constant placement of subgenus *Milecovirus* (MLeV), we can extend this pattern to the relationship ((((A, B) Γ) Δ) MLeV). This natural hierarchy within Coronaviridae is in principle congruent to the general pattern of their hosts: (((Mammals) Birds plus Mammals) Amphibia) (Table S1).
3. Subgenus *Hibecovirus* has been proposed as a sister taxon of the subgenus *Sarbecovirus* (reviewed in (*5*)). This proposal, which again was made on the basis of unrooted trees, also benefited from reexamination using the comprehensive rooted trees produced in this study. All four methods validate the sister relationship of the subgenera *Hibecovirus* and *Sarbecovirus* clade (Figure 1, Figures S1 - S5).
4. Because of our particular interest in regards to SARS-Cov-2, we would like to stress that all four methods have placed pangolin coronavirus as a sister of the clade {RaTG13 plus RmYN02 plus SARS-Cov-2} (Figure 1, Figures S1 - S5), confirming that bat

coronaviruses RaTG13 or RmYN02, but not the pangolin coronavirus, are indeed the closest relatives of SARS-Cov-2 (Figure 1, Figures S1 - S5). In other words, contrary to some recent suggestions (reviewed in *17*, *18*) the rooted trees produced herein confirm that either bat coronavirus RaTG13 *(21)* or RmYN02 *(24)*, but not the pangolin coronavirus, is an immediate sister of SARS-Cov-2 (Figure 1, Figures S1 - S5).

Differences in details within clade relationships do exist between the ML method and each of the additional cladistic methods newly applied to molecular sequence data of Coronaviridae. For example the MP, AC and 3TA trees (Figures S1-S3), but not ML trees (Figure S4), have placed Lucheng Rn rat coronavirus (genus *Alphacoronavirus*, subgenus *Luchacovirus*) as a sister taxon to the clade that contains all of the remaining members of genus *Alphacoronavirus*.

Similarly, both conventional phylogenetic methods (MP and ML) have defined bat coronavirus RmYN02 as a weakly supported sister of the SARS-Cov-2 (Figures S1, S4). However both Hennigian methods (3TA and AC) are, in contrast, placed bat coronavirus RaTG13, but not RmYN02, as a sister of this newly discovered coronavirus (Figures S2, S3).

All four analyses are initially based on a common molecular matrix (22,489 bp G-Block version of the 47-genomes alignment) (Supplementary Materials). Differences between analyses are likely a result of how each method deals with conflict to form the optimal trees. Nevertheless, the similarity between the tree topologies suggests that, regardless of method, many of the nodes are 'true' summaries of the data and that the data themselves are relatively noise-free (Figure 1, Figures S1 - S5).

b. On relationships and hosting of newly discovered *Sarbecovirus* SARS-Cov-2

Even from the elementary comparative point of view, it is clear that every Coronaviridae virus seems to be well defined and well separated from the others sometimes by hundreds or (more commonly) thousands of single nucleotide positions (SNPs) (Table S2, S3). For example, the minimal relationship {RmYN02 {RaTG13 plus SARS-Cov-2}} (Figures S2, S3), based on the 29,907 bp complete genomic alignment of three taxa (RmYN02, RaTG13 and SARS-Cov-2 (MN908947)), implies 1,329 informative SNPs (or, respectively, 1,329 3TSs) from the total 2,467 variable characters.

Thus, even closely related viruses from the *Sarbecovirus* clade (including the newly discovered SARS-Cov-2 and bat coronaviruses RaTG13 and RmYN02), are all remarkably different from one another from a comparative standpoint. The same is true for every relationship within Coronaviridae we have recovered in our analyses.

Such simple observations automatically exclude the possibility of the recombination origins of SARS-Cov-2 (reviewed in (*19*), see also (*8, 20*)) as well as other similar propositions. Based on the data available to date, including the comprehensive trees produced herein (Figure 1, Figures S1 - S5), the focus should shift to the static aspect of the problem.

The monophyly of all the genera of Coronaviridae, as well as all of its non-monotypic subgenera, and also the general relationship ((((A, B) Γ) Δ) within the subfamily Orthocoronavirinaeis, can also be demonstrated within the pure comparative analytical framework (Figures S2 and S3). Within the later, no member of *Sarbecovirus* clade is an ancestor of SARS-Cov-2. This static view may be critical in discussing the general simple pattern in hosting of SARS-Cov-2.

Recent studies of SARS-related coronaviruses have suggested that bats harbor close relatives to SARS-Cov-2 (for example (*21*)), and that pangolins may be natural hosts of this member of genus *Betacoronavirus* (*8, 18, 26*), leading to the hypothesis of animal to human transmission of SARS-Cov-2 ("The presence in pangolins of an RBD very similar to that of

SARS-CoV-2 means that we can infer this was also probably in the virus that jumped to humans" *(18)*).

However, the search of other hosts, as well as the related exotic ways of transmission of SARS-Cov-2 from these hypothetical hosts to humans, is based on a set of the complicated assumptions (for example, "The sequence similarity in the spike receptor binding domain between SARS-CoV-2 and a sequence from pangolin is probably due to an ancient intergenomic introgression..." *(26)*) and also ignores the simple possibility of the original human-based hosting of SARS-Cov-2.

In fact, the hosting of the viruses that are genetically related to SARS-Cov-2 by bats or pangolins is not, strictly speaking, an argument in favor of the animal hosting of SARS-Cov-2, especially because the latter virus had never been detected inside of animals such as bats. The clear possibility of the original human hosting of SARS-Cov-2, unfortunately, had not been discussed in scientific literature. However, the extremely high contagiousness of SARS-Cov-2 is almost improbable if the virus had been transmitted from animals to humans just several months ago. Alternatively, it is possible that a form of SARS-Cov-2 may have been existed in some parts of the human population before the current pandemic, and that, perhaps, the virus had been effectively suppressed by the human immune system for a sometime prior. Seven human coronaviruses have been identified to date (reviewed in (*22*)), yet at least three of these namely, Human coronaviruses OC43 and HKU1 from subgenus *Embecovirus* and NL63 from subgenus *Setracovirus* (Table S1) are globally distributed viruses where no animal hosts have been proposed.

With the information available to date, there is no direct evidence to suggest which hypothesis, animal to human transmission (either the recent or less recent one *(25, 26*)) or original human hosting, is true. The evidence only tells us how viruses are related within a hierarchical classification (Figure 1).

c. On taxonomy, naming, and nomenclature of the coronaviruses

Because the taxonomy and nomenclature of the viruses are still "under construction" (*13-16*), the names of the virus species frequently remain non-binary, even within the taxonomic statements of ICTV (*13-16*). Simultaneously, the current circumscriptions of the two largest genera of the Coronaviridae (*Alphacoronaviruses* and *Betacoronaviruses*) are very complicated. For example, current genus *Alphacoronavirus* currently circumscribes 12 subgenera and current genus *Betacoronavirus* circumscribes five subgenera. These two observations cause issues with a clear naming of viruses on the phylogenetic trees of Coronaviridae, as well as with the reading of these trees, especially by a non-specialist.

Here we resolved these issues by using the ICTV summaries (*13-16*) of the **subgeneric** names of Coronaviridae as the basic units of our notation where possible in the analyses and trees. In other words, whenever possible, abbreviations and trivial names were avoided to improve clarity. The recently discovered SARS-Cov-2 virus and a few related viruses are the exceptions (Table S1).

Reconsideration of the current circumscriptions of Coronaviridae genera may provide a simpler and more informative taxonomic system of both naming and nomenclature. Accepting traditional genera of the family at the rank of a tribe and, simultaneously, the current subgenera (all of which are monophyletic (Figure 1, Figures S1 - S5)) at the generic rank seems to be the easiest heuristic way to incorporate the Linnaean principles of the binary nomenclature right to the classification of the family. For example, the current member of the subgenus *Sarbecovirus* (current genus *Betacoronavirus*) virus SARS-Cov-2 may be easily named ***Sarbecovirus* species, abbreviation: sp. (e.g.: *Sarbecovirus* sp.)**, where "species" (sp.) is any available epithet. Numerous currently discovered variants of it (summarized in (*6*)) can be *in principle* established

as forms of varieties of the same species. The trivial name “severe acute respiratory syndrome-related coronavirus 2” and correspondent abbreviation “SARS-Cov-2” can be simply listed in the description of the species.

It is also critical to consistently involve the type method of biological taxonomy to the taxonomy of the viruses. As a simple example, we would like to suggest that the ICTV approved GenBank Accession number (ideally the reference to the whole genome of the virus) can be used as a nomenclature type of any virus species. For example the GenBank Accession number MN908947 can be treated as a nomenclature type of newly described *Sarbecovirus* (SARS-Cov-2). Such a number implies the name/abbreviation of the biological isolate as well as other useful information. The higher nomenclature categories (tribes, genera, families etc.) can be typified by the names of the species (genera etc.), exactly in a manner of the plant or animal names. The nomenclature classification of plants and animals has developed over hundreds of years, and as such is robust and well tested. Adopting the Linnaean binary nomenclature for viruses will increase the universality of the system, and thereby lead to more consistent information content and information exchange.

We believe that such simple recommendations are fully consistent with the principles of the future binary nomenclature of viruses, currently summarized by Siddell *et al*. (*15, 16*) and others (*23*) and may be useful for the different families of viruses. Such a nomenclature system would help scientists find information about a particular taxon easily and quickly, which is a high priority when accurate and timely identification is required during pandemic outbreaks.

Hierarchial classification of family Coronaviridae

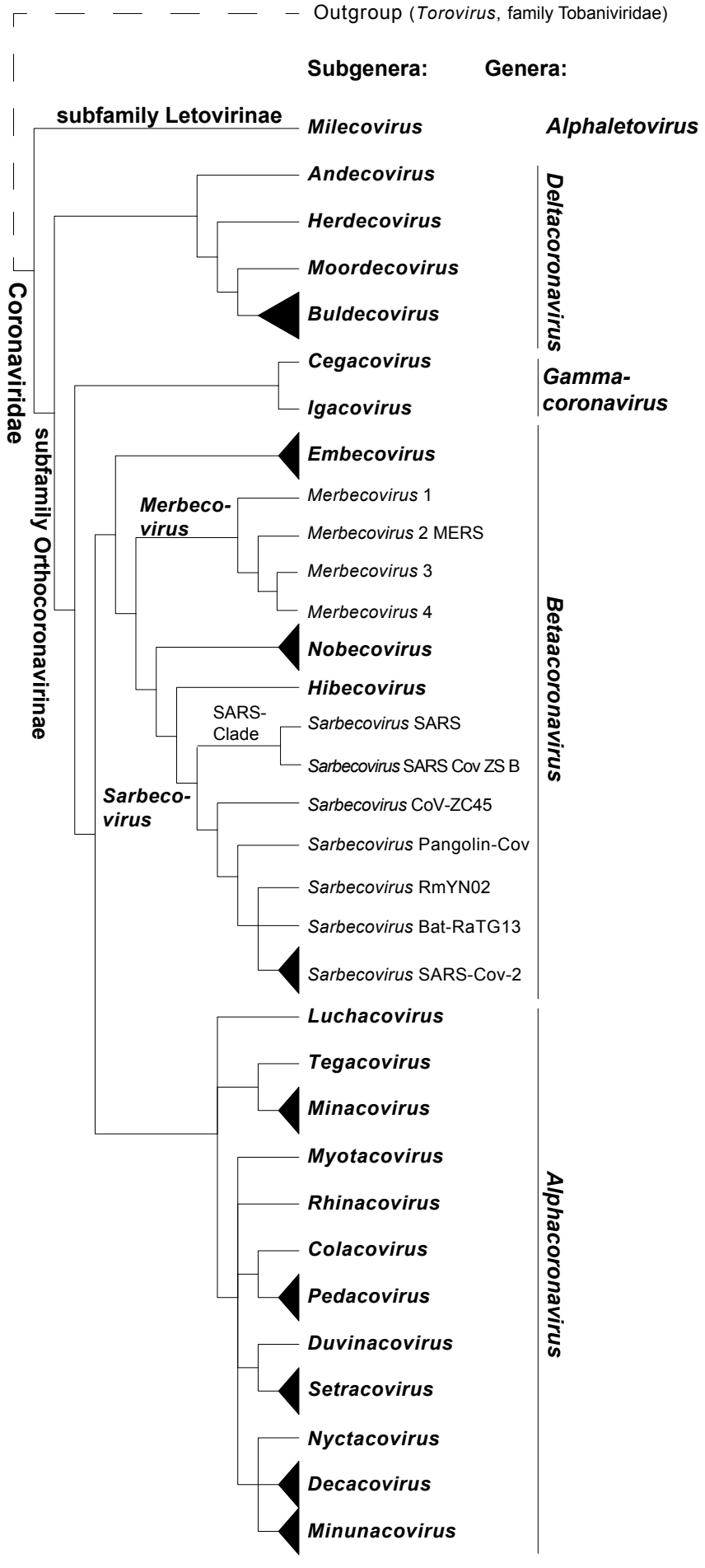

Figure 1

Figure 1. Hierarchical classification of the coronaviruses (Riboviria, Nidovirales, Coronaviridae), established as a simplified Strict Consensus of four trees, produced by three different methods of cladistic analysis as well as by Maximum Likelihood method using G-block version of the genomic alignment of Coronaviridae + *Torovirus*. See Methods (Supplementary Materials), Figures S1 - S5 and Table S1 for more detail including the tree node support values. Figure was drawn with special attention to the taxonomic placements of highly pathogenic Middle East respiratory syndrome-related coronavirus (MERS-Cov) and severe acute respiratory syndrome-related coronaviruses 1 and 2 (SARS-CoV and SARS-Cov-2).

Acknowledgments

Author contributions: E.V.M.: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. N.E.M.: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Writing - review & editing, Visualization. M.C.E.: Conceptualization, Methodology, Formal Analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision. D.M.W.: Conceptualization, Methodology, Formal Analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision. M.L.T.: Conceptualization, Investigation, Data Curation, Writing - original draft, Writing - review & editing, Supervision. **Competing interests:** authors declare no competing interests. **Data and materials availability:** "All data is available in the main text or the supplementary materials."

References

1. N. J. Maclachlan, E. J. Dubovi, S. W. Barthold, D. F. Swayne, J. R. Winton, *Fenner's Veterinary Virology: Fifth edition*. (Elsevier Inc, Amsterdam, 2016).

2. K. Bukhari *et al.*, Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. *Virology* **524**, 160 (Nov, 2018).

3. A. E. Gorbalenya *et al.*, The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* **5**, 536 (Apr, 2020).

4. R. Tokarz *et al.*, Discovery of a novel nidovirus in cattle with respiratory disease. *J Gen Virol* **96**, 2188 (Aug, 2015).

5. A. C. P. Wong, X. Li, S. K. P. Lau, P. C. Y. Woo, Global Epidemiology of Bat Coronaviruses. *Viruses* **11**, (Feb 20, 2019).

6. D. S. Candido *et al.*, Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*, (Jul 23, 2020).

7. F. Ferron, H. J. Debat, A. Shannon, E. Decroly, B. Canard, A N7-guanine RNA cap methyltransferase signature-sequence as a genetic marker of large genome, non-mammalian Tobaniviridae. *NAR Genomics and Bioinformatics* **2**, lqz022 (Mar, 2020).

8. P. Liu *et al.*, Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathogens* **16**, e1008421 (May, 2020).

9. R. Lu *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565 (Feb 22, 2020).

10. F. Wu *et al.*, Author Correction: A new coronavirus associated with human respiratory disease in China. *Nature* **580**, E7 (Apr, 2020).

11. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265 (Mar, 2020).

12. I. J. Kitching *et al.*, *Cladistics: The Theory and Practice of Parsimony Analysis*. (Oxford University Press, 1998).

13. M. J. Adams *et al.*, 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Archives of virology* **162**, 1441 (May, 2017).

14. P. J. Walker *et al.*, Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Archives of virology* **164**, 2417 (Sep, 2019).

15. S. G. Siddell *et al.*, Correction to: Binomial nomenclature for virus species: a consultation. *Archives of virology* **165**, 1263 (May, 2020).

16. S. G. Siddell *et al.*, Binomial nomenclature for virus species: a consultation. *Archives of virology* **165**, 519 (Feb, 2020).

17. S.-L. Liu, L. J. Saif, S. R. Weiss, L. Su, No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. *Emerging Microbes & Infections* **9**, 505 (2020).

18. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nature medicine* **26**, 450 (Apr, 2020).

19. X. Li *et al.*, Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Science Advances* **6**, eabb9153 (2020).

20. M. F. Boni *et al.*, Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, (Jul 28, 2020).

21. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270 (Mar, 2020).

22. D. X. Liu, J. Q. Liang, T. S. Fung, Human Coronavirus-229E, -OC43, -NL63, and -HKU1. *Reference Module in Life Sciences*, B978 (2020).

23. International Committee on Taxonomy of Viruses Executive Committee, The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology* **5**, 668 (May, 2020).

24. H. Zhou *et al*., A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Current Biology* **30**, (2020).

25. L. Pipes *et al.*, Assessing uncertainty in the rooting of the SARS-CoV-2 phylogeny. *bioRxiv*, (2020).

26. S. M. Chaw *et al*., The origin and underlying driving forces of the SARS-CoV-2 outbreak. *Journal of Biomedical Science* **27**:73. (2020).

Supplementary Materials for

On Classification and Taxonomy of Coronaviruses (Riboviria, Nidovirales, Coronaviridae) with special focus on severe acute respiratory syndrome-related coronavirus 2 (SARS-Cov-2)

Evgeny V. Mavrodiev, Melinda L. Tursky, Nicholas E. Mavrodiev, Malte C. Ebach, David M. Williams.

Correspondence to: evgeny@ufl.edu

This PDF file includes:

Materials and Methods
Supplementary Text
Figures S1 to S5
Captions for Figures S1 to S5

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S3
Captions for Tables S1 to S3

Materials and Methods

a. Taxonomic sampling of the study

Thirty nine ICTV-approved genomes (*1- 4*) of all species of Coronaviridae have been used in this study (Table S1).

Additionally to these 39 genomes, we also included along with the final alignments

a. the published genome MN908947 of SARS-Cov-2(*5, 6*) (Table S1);

b. an unpublished genome of the same virus species (MN988713, Tao *et al*., unpublished) (Table S1);

c. the genome of bat SARS-like coronavirus (bat-SL-CoV-ZC45; MG772933) (Table S1) that has been previously used in other comparisons to SARS-Cov-2 (for instance (*5*-7 and others*)*);

d. - f. bat coronavirus RaTG13 (MN996532 (*8*) (Table S1)), pangolin coronavirus (MT121216 (*9*)) (Table S1) and RmYN02 (GISAID: EPI_ISL_412977) (*32*) (Table S1)) as they have previously been proposed as the closest known relatives of SARS-Cov-2 (reviewed in (*9, 10, 32, 33*));

g. an additional SARS related genome ZS-B (AY394996) (Table S1).

The genome of recently discovered *Microhyla letovirus* 1 or MLeV virus (subgenus *Milecovirus*) (*11*) was also included with the alignments and analyses. This genome was available upon courtesy request from Prof. B. W. Neuman (Texas A&M University-Texarkana, TX, US) (Table S1).

Based on the summary of Maclachlan *et al*. (*12*), genus *Torovirus* (family Tobaniviridae, order Nidovirales) was assumed as the best outgroup taxon of Coronaviridae, and ICTV-approved genome of *Torovirus* (AY427798) (Table S1) has been selected for this study.

The names of the major obtained relationships on all trees (Figure S1–S5) excluding the names of the families and subfamilies, are written in italics due to the strong congruence with different taxonomic entities. Depending on the context, the names of the clades are provided in regular and/or curved brackets.

As the utility of phylogenetic trees depends on their clarity, the use of abbreviations and trivial names of viruses has been avoided whenever possible.

b. Matrices and trees

All genomic alignments have been performed using MAFFT (*13, 14, 15*) following FFT-NS-I strategy with the command: mafft --inputorder --adjustdirection --anysymbol --kimura 1 --maxiterate 1000 --6merpair input.

Poorly aligned positions have been removed from

1. the genomic alignment of Coronaviridae plus *Torovirus* and
2. the genomic alignment of subfamily Orthocoronavirinae with no outgroups included

using the program G-block (*16*) as implemented in SeaView (*17*) under the conditions of "less stringent" strategy of the algorithm (*16, 17*).

The G-block version of the genomic alignment of Coronaviridae plus *Torovirus* was also established as a binary matrix using simple "presence – absence" coding (reviewed in (*18*)) with the future inclusion of the "all-plesiomorphic" ("all-zero") artificial taxon. In short, G-block based genomic alignment of the family Coronaviridae plus *Torovirus* was rewritten as a binary (01) matrix, where "zero" means "the absence of a nucleotide in this particular position of the alignment", and "one" means "the presence of the nucleotide in this particular position of the same alignment". For example, if the character-state of the character number 253 is equal to A (Adenine), than this can be written as "1000", where "1" means "the A is present in position 253", and "0" indicates that U(T), G and C are simultaneously absent on the same position.

Assuming that the "absence of the nucleotide" (the character-state "zero") is a plesiomorphic character-state, we can add to the binary matrix "all-plesiomorphic" or "all-zeros" outgroup. The binary matrix with an "all-zeros" outgroup added was later used as an input into the script Forester v. 1.0 (*19*) following the command ruby trees.rb path_to_matrix_file (*19*) with future selection of the "Additional" forest of the maximal trees (relationships) (*19*) for Average Consensus (AC) analysis (*19-21*).

Manipulations with either the molecular or binary matrices and the tree-files have been performed with Mesquite v. 3.51 (*22*), PAUP* v. 4.0a (*23*) and FigTree v. 1.4.2 (*24*).

3. Analyses

The G-block version of the molecular alignment of Coronaviridae plus *Torovirus* was analysed by the standard Maximum Parsimony (MP) approach (Fitch Parsimony, reviewed in (*18*)), and by the three-taxon statement analysis (3TA) with fractional weighting (reviewed in (*18, 27, 28*) and implemented in Mavrodiev and Madorsky (*29*)).

Following the logic of Williams-Siebert (WS) representation of the unordered multistate data (reviewed in (*29*)) the three-taxon statement (3TS) permutations of the G-block-based alignment of Coronaviridae plus *Torovirus* were conducted with TAXODIUM version 1.2 using the command: taxodium.exe input_file_name.csv –idna –ob –og –fw –nex *(29)* taking values of the operational outgroup as equal to the values of *Torovirus*.
All MP analyses have been performed in PAUP* (*23*) as in Mavrodiev *et al*. and Mavrodiev and Madorsky (*19, 29*). The resulted most parsimonious tree resulted standard MP analysis was *a posteriori* rooted relative to *Torovirus*.

The AC (*20, 21*) of the array of maximal trees was calculated using the program Clann version 4.1.5 (*25, 26*) as described in Mavrodiev *et al*. (*19*). The distance optimality criterion for the AC analysis was specified as a "distance with non-weighted least squares" (*20, 21, 23, 25*).

Following conclusions of Zhou, X. *et al*.(*30*), the Maximum Likelihood (ML) analysis of G-block alignment of Coronaviridae plus *Torovirus* was conducted with W-IQ-TREE (*31*) with implemented automatic model selection procedure (*31*). The resulted most probable tree was *a posteriori* rooted relative to *Torovirus*.

The MP Bootstrap support (BS) values have been calculated as described in (*29*) and *(31)*. In the case of the ML analysis, the aLRT support values have been calculated instead of the ML BS supports, as implemented W-IQ-TREE (*31*).

The simplest "total" character differences between the G-block modified aligned genome sequences of subfamily Orthocoronavirinae (Table S2), as well as between three aligned genomes of bat coronaviruses RmYN02, RaTG13 and severe acute respiratory

syndrome-related coronavirus 2 (SARS-Cov-2) (MN908947) with no aligned positions excluded (Table S3), was calculated in PAUP*(*23*) under the default options.

d. Supplementary Text

The genomic alignment of Coronaviridae + *Torovirus* (outgroup) consists of 52,990 molecular characters (base pairs or bp), the G-block version of this alignment is of 22,489 characters with 19,550 of those are parsimony-informative. This alignment was the target of future MP, 3TA, AC and ML analyses.

The genomic alignment of subfamily Orthocoronavirinae with no outgrop included is of 49,881 bp. The G-block version of this alignment consists of 23,431 bp. Later, the alignment has been used only to calculate the simplest pairwise distances ("total character differences") between all of the members of the subfamily included in the analyses (Table S2). The genomic alignment of bat coronaviruses RmYN02, RaTG13 and severe acute respiratory syndrome-related coronavirus 2 (SARS-Cov-2) (MN908947) is of 29,907 molecular characters. This alignment with no characters excluded was used to calculate the total character differences (Table S3) between newly discovered SARS-Cov-2 and two of its closest relatives. Such a measure is the simplest expression of the pairwise distance between aligned molecular sequences and indicates solely the total number of different single nucleotide positions (or SNPs) between them. For example, the number 1138 in the Table S3 means that the aligned genomes of human coronavirus SARS Cov 2 (accession "a") and the bat coronavirus RaTG13 are different from each other by 1138 SNPs. For instance, in position # 37 of the same molecular alignment the value of SARS Cov 2 is equal to "C" and the value of RaTG13 is equal to "G". The total number of such SNPs equals 1138.

The standard MP analysis of the 22,489 bp G-block alignment resulted in the single most parsimonious tree of 208,176 steps (CI = 0.2505, RI = 0.4954) (Figure S1).

The 3TS representation of the same 22,489 bp G-block alignment resulted 39,621,820 3TSs (binary characters), all parsimony-informative and fractionally weighted, with the most parsimonious fit of 20,786,459.7424 steps (RI = 0.5706) (Figure S2).

The presence-absence re-coding of the genomic alignment of Coronaviridae + *Torovirus* resulted in a matrix of 73,258 binary characters from which 65,435 characters can be established as a rooted trees (maximal relationships). The AC analyses of the forest of these 65,435 relationships resulted in a single AC tree of the score 0.00911 (Figure S3).

For ML analysis of the 22,489 bp G-block matrix of Coronaviridae and outgroup (*Torovirus*), GTR+F+R10 model has been automatically selected by W-IQ-TREE as a best-fit model based on either corrected and non-corrected Akaike Information Criteria, as well as on Bayesian Information Criterion. The resulted single most probable (ML) tree has the best score (log likelihood) equal to -766940.2344 (Figure S4).

Standard Maximum Parsimony analysis

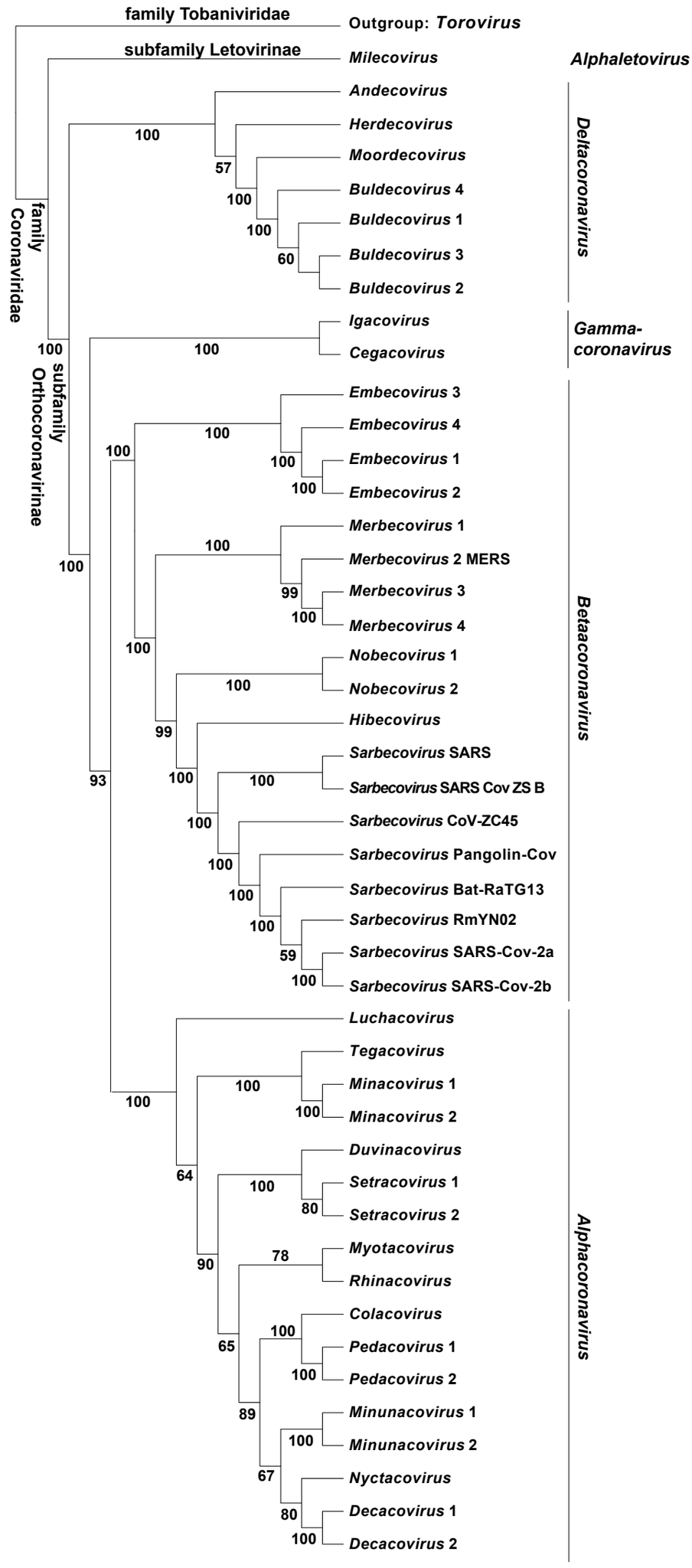

Figure S1

Three-taxon statment analysis (WS representation, fracrional weighting)

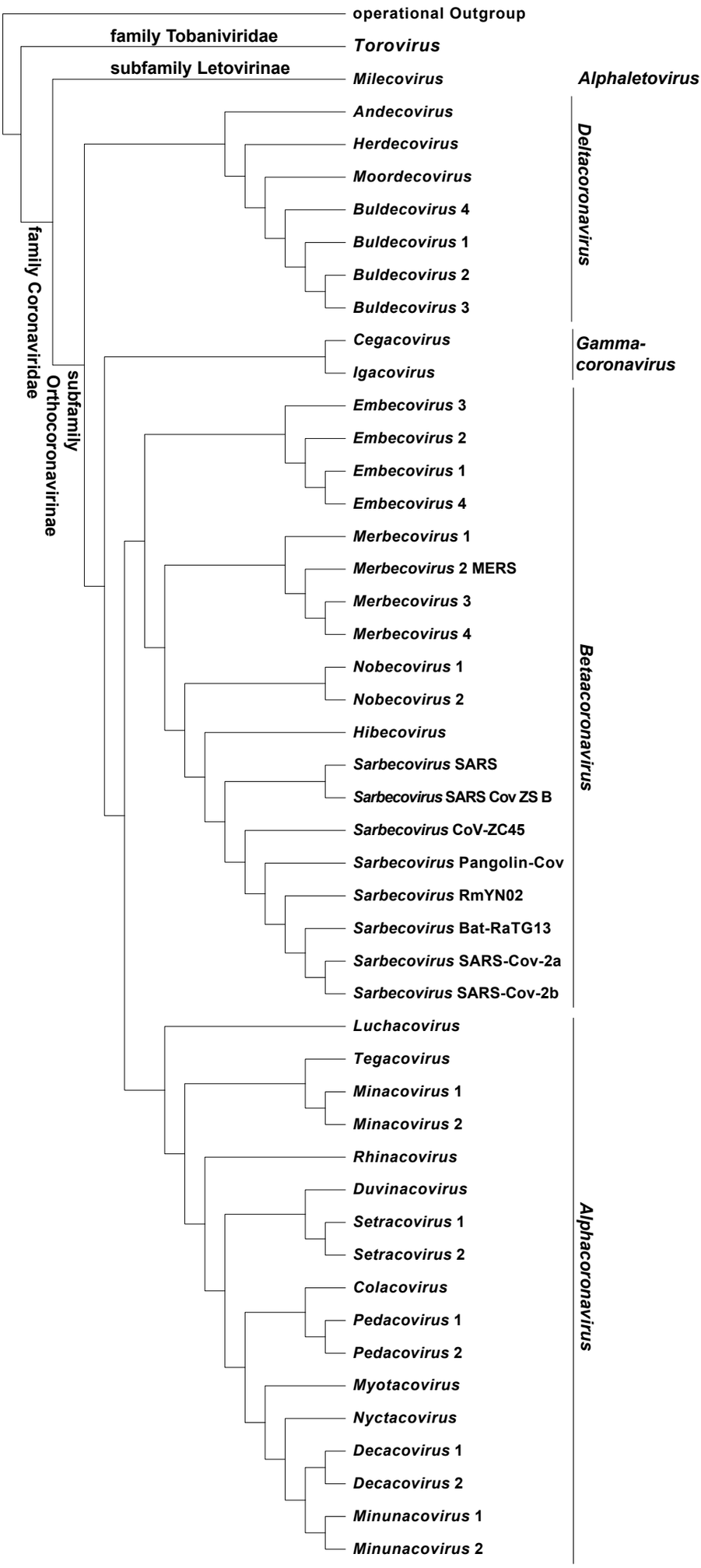

Figure S2

Average Consensus analysis

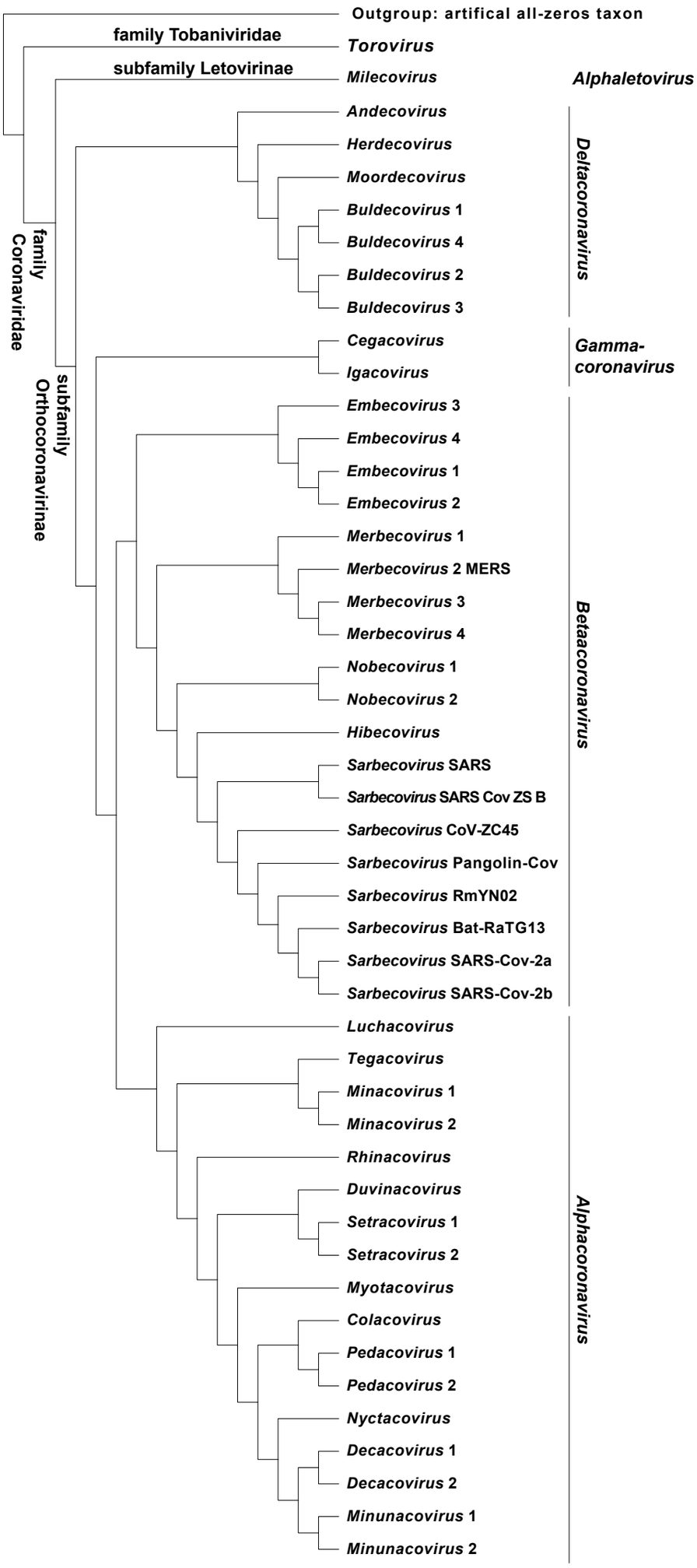

Figure S3

Maximum Likelihood analysis

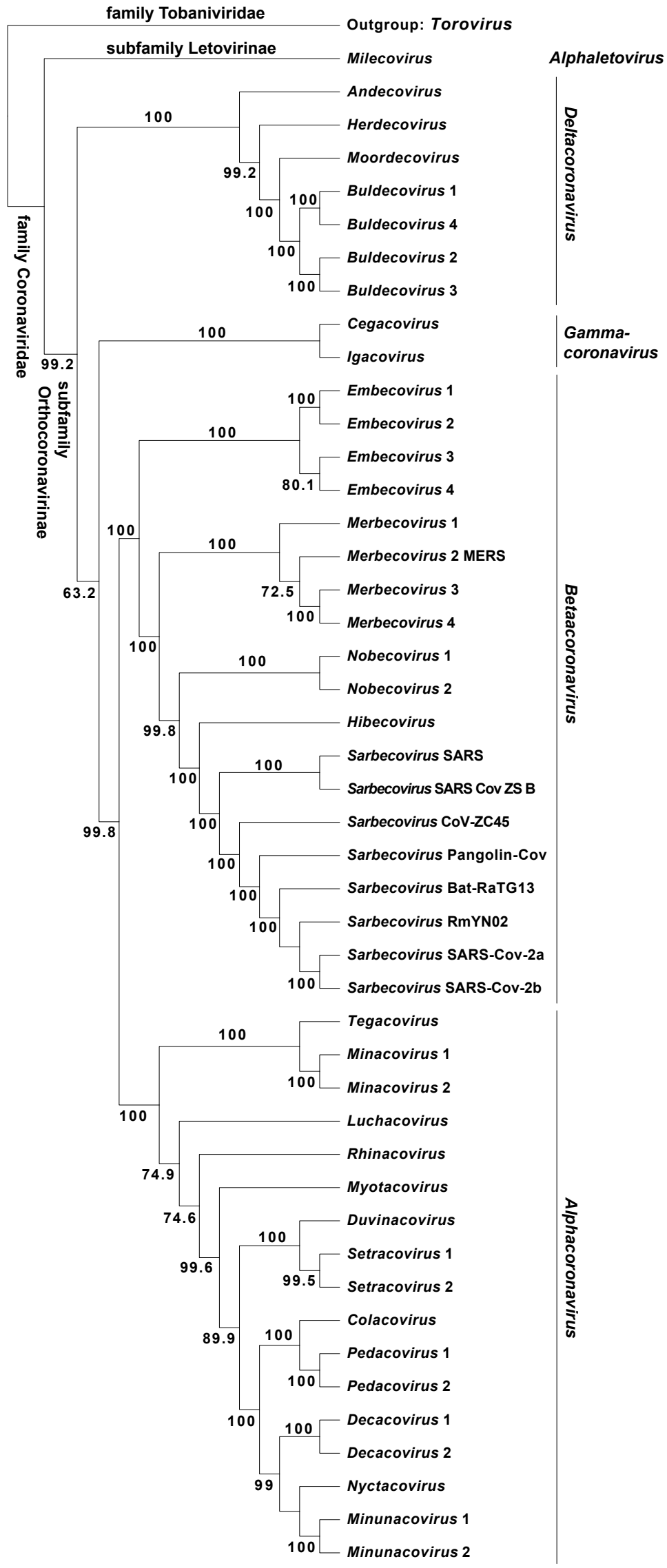

Figure S4

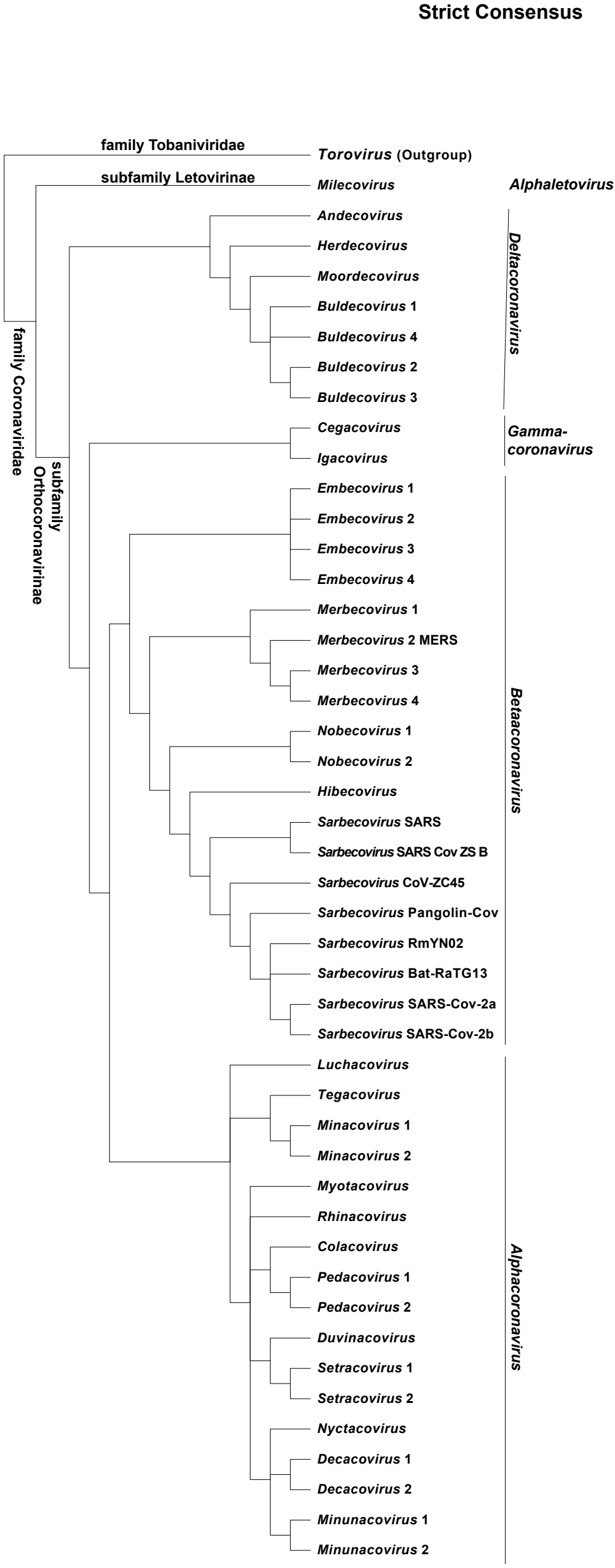
Strict Consensus
family Tobaniviridae
Torovirus (Outgroup)
subfamily Letovirinae
Milecovirus
Alphaletovirus
Andecovirus
Herdecovirus
Moordecovirus
Buldecovirus 1
Buldecovirus 4
Buldecovirus 2
Buldecovirus 3
Deltacoronavirus
Cegacovirus
Igacovirus
Gamma-coronavirus
Embecovirus 1
Embecovirus 2
Embecovirus 3
Embecovirus 4
Merbecovirus 1
Merbecovirus 2 MERS
Merbecovirus 3
Merbecovirus 4
Nobecovirus 1
Nobecovirus 2
Hibecovirus
Sarbecovirus SARS
Sarbecovirus SARS Cov ZS B
Sarbecovirus CoV-ZC45
Sarbecovirus Pangolin-Cov
Sarbecovirus RmYN02
Sarbecovirus Bat-RaTG13
Sarbecovirus SARS-Cov-2a
Sarbecovirus SARS-Cov-2b
Betaacoronavirus
Luchacovirus
Tegacovirus
Minacovirus 1
Minacovirus 2
Myotacovirus
Rhinacovirus
Colacovirus
Pedacovirus 1
Pedacovirus 2
Duvinacovirus
Setracovirus 1
Setracovirus 2
Nyctacovirus
Decacovirus 1
Decacovirus 2
Minunacovirus 1
Minunacovirus 2
Alphacoronavirus
family Coronaviridae
subfamily Orthocoronavirinae

Figure S5

Captions for Figures S1 to S5

Figure S1.

Single most parsimonious tree recovered from the standard MP analysis (Fitch Parsimony) of the 22,489 bp genomic alignment of Coronaviridae + *Torovirus*. Tree was *a posteriori* rooted relative to *Torovirus*.

Figure S2.

Most parsimonious hierarchy of patterns recovered from MP analysis (Wagner parsimony) of 39,621,820 3TS WS representation of the 22,489 bp genomic alignment of Coronaviridae + *Torovirus*. The values of the operational outgroup were fixed as values of *Torovirus*.

Figure S3.

The average consensus tree from the analysis of the forest of 65,435 rooted trees (maximal relationships) derived from the "presence-absence" representation of the original 22,489 bp genomic alignment of Coronaviridae + *Torovirus*.

Figure S4.

Most probable tree recovered from the ML analysis of the 22,489 bp genomic alignment of Coronaviridae + *Torovirus*. Tree was *a posteriori* rooted relative to *Torovirus*.

Figure S5.

Strict Consensus of four trees produced by three different methods of cladistic analysis as well as by Maximum Likelihood method using modified genomic alignment of Coronaviridae + *Torovirus* (Figures S1 – S4).

Table S1: Taxonomic sampling of the study and related data

Realm	Order	Family	Subfamily	Genus	Subgenus	Species	Virus name	Abbreviation and/or the name of the isolate	Name in the Trees	Host	Accession number	Source
Riboviria	*Nidovirales*	*Coronaviridae*	*Letovirinae*	*Alphaletovirus*	*Milecovirus*	*Microhyla letovirus* 1	Microhyla letovirus 1	MLeV	*Milecovirus*	Amphibia	n/a	Courtesy of Prof. B. W. Neuman
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Colacovirus*	Bat coronavirus CDPHE15	bat coronavirus CDPHE15	BtCoV CDPHE15	*Colacovirus*	Mammals	KF430219	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Decacovirus*	Bat coronavirus HKU10	rousettus bat coronavirus HKU10	BtCoV HKU10	*Decacovirus*-1	Mammals	JQ989270	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Decacovirus*	Rhinolophus ferrumequinum alphacoronavirus HuB-2013	BtRf-AlphaCoV/HuB2013	BtRf-AlphaCoV	*Decacovirus*-2	Mammals	KJ473807	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Duvinacovirus*	Human coronavirus 229E	human coronavirus 229E	HCoV-229E	*Duvinacovirus*	Mammals	AF304460	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Luchacovirus*	Lucheng Rn rat coronavirus	Lucheng Rn rat coronavirus	LRNV	*Luchacovirus*	Mammals	KF294380	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Minacovirus*	Ferret coronavirus	ferret coronavirus	FRCoV	*Minacovirus*-1	Mammals	LC119077	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Minacovirus*	Mink coronavirus 1	mink coronavirus	MCoV	*Minacovirus*-2	Mammals	HM245925	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Minunacovirus*	Miniopterus bat coronavirus 1	miniopterus bat coronavirus 1	Mi-BatCoV-1A	*Minunacovirus*-1	Mammals	EU420138	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Minunacovirus*	Miniopterus bat coronavirus HKU8	miniopterus bat coronavirus HKU8	Mi-BatCoV-HKU8	*Minunacovirus*-2	Mammals	EU420139	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Myotacovirus*	Myotis ricketti alphacoronavirus Sax-2011	BtMr-AlphaCoV/SAX2011	BtMr-AlphaCoV	*Myotacovirus*	Mammals	KJ473806	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Nyctacovirus*	Nyctalus velutinus alphacoronavirus SC-2013	BtNv-AlphaCoV/SC2013	BtNv-AlphaCoV	*Nyctacovirus*	Mammals	KJ473809	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Pedacovirus*	Porcine epidemic diarrhea virus	porcine epidemic diarrhea virus	PEDV	*Pedacovirus*-1	Mammals	AF353511	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Pedacovirus*	Scotophilus bat coronavirus 512	scotophilus bat coronavirus 512	Sc-BatCoV-512	*Pedacovirus*-2	Mammals	DQ648858	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Rhinacovirus*	Rhinolophus bat coronavirus HKU2	rhinolophus bat coronavirus HKU2	Rh-BatCoV-HKU2	*Rhinacovirus*	Mammals	EF203064	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Setracovirus*	Human coronavirus NL63	human coronavirus NL63	HCoV-NL63	*Setracovirus*-1	Mammals	AY567487	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Setracovirus*	NL63-related bat coronavirus strain BtKYNL63-9b	NL63-related bat coronavirus	BtKYNL63	*Setracovirus*-2	Mammals	KY073745	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Alphacoronavirus*	*Tegacovirus*	Alphacoronavirus 1	transmissible gastroenteritis virus	TGEV	*Tegacovirus*	Mammals	AJ271965	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Embecovirus*	Betacoronavirus 1	human coronavirus OC43	HCoV-OC43	*Embecovirus*-1	Mammals	AY585228	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Embecovirus*	China Rattus coronavirus HKU24	betacoronavirus HKU24	ChRCoV-HKU24	*Embecovirus*-2	Mammals	KM349742	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Embecovirus*	Human coronavirus HKU1	human coronavirus HKU1	HCoV-HKU1	*Embecovirus*-3	Mammals	AY597011	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Embecovirus*	Murine coronavirus	murine hepatitis virus	MHV	*Embecovirus*-4	Mammals	AY700211	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Hibecovirus*	Bat Hp-betacoronavirus Zhejiang2013	bat Hp-betacoronavirus/Zhejiang2013	Bat-Hp-BetaCoV	*Hibecovirus*	Mammals	KF636752	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Merbecovirus*	Hedgehog coronavirus 1	hedgehog coronavirus 1	EriCoV	*Merbecovirus*-1	Mammals	KC545383	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Merbecovirus*	Middle East respiratory syndrome-related coronavirus	Middle East respiratory syndrome-related coronavirus	MERS-CoV	*Merbecovirus*-2-MERS	Mammals	JX869059	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Merbecovirus*	Pipistrellus bat coronavirus HKU5	pipistrellus bat coronavirus HKU5	Pi-BatCoV-HKU5	*Merbecovirus*-3	Mammals	EF065509	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Merbecovirus*	Tylonycteris bat coronavirus HKU4	tylonycteris bat coronavirus HKU4	Ty-BatCoV-HKU4	*Merbecovirus*-4	Mammals	EF065505	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Nobecovirus*	Rousettus bat coronavirus GCCDC1	rousettus bat coronavirus	Ro-BatCoV-GCCDC1	*Nobecovirus*-1	Mammals	KU762338	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Nobecovirus*	Rousettus bat coronavirus HKU9	rousettus bat coronavirus HKU9	Ro-BatCoV-HKU9	*Nobecovirus*-2	Mammals	EF065513	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*	Severe acute respiratory syndrome-related coronavirus	severe acute respiratory syndrome-related coronavirus	SARS-CoV	*Sarbecovirus-SARS*	Mammals	AY274119	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		severe acute respiratory syndrome-related coronavirus 2	SARS-CoV2	*Sarbecovirus* SARS Cov 2a	Mammals	MN908947	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		severe acute respiratory syndrome-related coronavirus 2	SARS-CoV2	*Sarbecovirus* SARS Cov 2b	Mammals	MN988713	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		bat coronavirus RaTG13	Bat-Cov-RaTG13	*Sarbecovirus* bat-RaTG13	Mammals	MN996532	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		pangolin coronavirus	Pangolin-Cov	*Sarbecovirus* Pangolin-CoV	Mammals	MT121216	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		SARS coronavirus ZS B	SARS-CoV ZS B	*Sarbecovirus* SARS Cov ZS B	Mammals	AY394996	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		bat coronavirus RmYN02	RmYN02	*Sarbecovirus* RmYN02	Mammals	EPI_ISL_412977	GISAID
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Betacoronavirus*	*Sarbecovirus*		bat SARS-like coronavirus	Bat-SL-CoV-ZC45	*Sarbecovirus* CoV-ZC45	Mammals	MG772933	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Andecovirus*	Wigeon coronavirus HKU20	wigeon coronavirus HKU20	WiCoV-HKU20	*Andecovirus*	Birds	JQ065048	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Buldecovirus*	Bulbul coronavirus HKU11	bulbul coronavirus HKU11	BulCV-HKU11	*Buldecovirus*-1	Birds	FJ376619	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Buldecovirus*	Coronavirus HKU15	porcine coronavirus HKU15	PoCoV-HKU15	*Buldecovirus*-2	Mammals	JQ065043	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Buldecovirus*	Munia coronavirus HKU13	munia coronavirus HKU13	MunCV-HKU13	*Buldecovirus*-3	Birds	FJ376622	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Buldecovirus*	White-eye coronavirus HKU16	white-eye coronavirus HKU16	WECoV-HKU16	*Buldecovirus*-4	Birds	JQ065044	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Herdecovirus*	Night heron coronavirus HKU19	night heron coronavirus HKU19	NHCoV-HKU19	*Herdecovirus*	Birds	JQ065047	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Deltacoronavirus*	*Moordecovirus*	Common moorhen coronavirus HKU21	common moorhen coronavirus HKU21	CMCoV-HKU21	*Moordecovirus*	Birds	JQ065049	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Gammacoronavirus*	*Cegacovirus*	Beluga whale coronavirus SW1	beluga whale coronavirus	BWCoV	*Cegacovirus*	Mammals	EU111742	GenBank
Riboviria	*Nidovirales*	*Coronaviridae*	*Orthocoronavirinae*	*Gammacoronavirus*	*Igacovirus*	Avian coronavirus	infectious bronchitis virus	IBV	*Igacovirus*	Birds	M95169	GenBank
Riboviria	*Nidovirales*	*Tobaniviridae*		*Torovirus*	*Renitovirus*	Bovine torovirus	bovine torovirus	BRV	*Torovirus*/Outgroup	Mammals	AY427798	GenBank

Table S2: Total character differences between the aligned genomes of the all coronaviruses from the subfamily Orthocoronavirinae based on the G-block version of the genomic alignment

# of taxon	Genus	Subgenus\Name in the Tree	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	*Alphacoronavirus*	*Colacovirus*	-																																												
2	*Alphacoronavirus*	*Decacovirus* 1	7709	-																																											
3	*Alphacoronavirus*	*Decacovirus* 2	7669	6017	-																																										
4	*Alphacoronavirus*	*Duvinacovirus*	8000	7773	7841	-																																									
5	*Alphacoronavirus*	*Luchacovirus*	9534	9417	9499	9556	-																																								
6	*Alphacoronavirus*	*Minacovirus* 1	8943	8831	8850	8756	9737	-																																							
7	*Alphacoronavirus*	*Minacovirus* 2	8862	8766	8766	8716	9573	5158	-																																						
8	*Alphacoronavirus*	*Minunacovirus* 1	7941	7472	7342	7989	9536	8894	8766	-																																					
9	*Alphacoronavirus*	*Minunacovirus*	7778	7422	7397	8043	9512	8928	8912	6584	-																																				
10	*Alphacoronavirus*	*Myotacovirus*	7961	7809	7843	8049	9469	9007	8940	7864	8035	-																																			
11	*Alphacoronavirus*	*Nyctacovirus*	8021	7688	7681	8216	9612	9108	8988	7728	7785	8116	-																																		
12	*Alphacoronavirus*	*Pedacovirus* 1	7214	7646	7640	8159	9724	9011	8954	7870	7760	8044	7916	-																																	
13	*Alphacoronavirus*	*Pedacovirus* 2	7248	7697	7749	8259	9591	9024	8994	7893	7849	8162	8091	6649	-																																
14	*Alphacoronavirus*	*Rhinacovirus*	8282	8050	8145	8154	9027	8921	8826	8103	8049	8102	8278	8391	8415	-																															
15	*Alphacoronavirus*	*Setracovirus* 1	7881	7569	7608	6875	9408	8643	8600	7669	7790	7898	8020	8019	7945	7896	-																														
16	*Alphacoronavirus*	*Setracovirus* 2	8002	7787	7831	7260	9651	8938	8880	8003	7964	8016	8153	8159	8094	8183	6679	-																													
17	*Alphacoronavirus*	*Tegacovirus*	9003	8942	8786	8872	9658	7070	7078	8973	9021	9003	9161	9119	9067	9008	8831	9013	-																													
18	*Betacoronavirus*	*Embecovirus* 1	10784	10812	10891	10836	10923	10900	10747	10844	10946	11000	10955	11122	11046	10713	10609	11036	10910	-																													
19	*Betacoronavirus*	*Embecovirus* 2	11184	11157	11222	11204	11193	11172	11182	11173	11219	11316	11226	11319	11262	11083	11006	11337	11214	5230	-																											
20	*Betacoronavirus*	*Embecovirus* 3	10715	10619	10702	10656	10818	10685	10593	10594	10842	10884	10828	10941	10821	10506	10267	10928	10710	5437	6220	-																										
21	*Betacoronavirus*	*Embecovirus* 4	11077	11158	11193	11178	11210	11136	11068	11224	11206	11203	11198	11289	11192	11042	11066	11221	11250	5239	5871	5570	-																									
22	*Betacoronavirus*	*Hibecovirus*	11130	11170	11088	11102	11364	11082	11120	11118	11253	11258	11268	11337	11205	11062	11058	11315	11098	10368	10605	10399	10565	-																								
23	*Betacoronavirus*	*Merbecovirus* 1	11005	10972	10953	10991	11216	11001	10982	11049	11022	11123	11163	11181	11073	10766	10784	11215	11061	10106	10390	9959	10480	10002	-																							
24	*Betacoronavirus*	*Merbecovirus* 2 MERS	11162	11173	11038	11118	11308	11141	11135	11172	11128	11300	11153	11322	11218	10981	10990	11194	11117	10300	10569	10177	10583	9950	6108	-																						
25	*Betacoronavirus*	*Merbecovirus* 3	11203	11277	11208	11360	11428	11242	11231	11385	11284	11390	11322	11356	11388	11252	11246	11448	11263	10550	10597	10556	10799	10132	6768	6001	-																					
26	*Betacoronavirus*	*Merbecovirus* 4	10921	10897	10860	10955	11054	10937	10961	10969	11003	11081	10984	11131	11071	10855	10675	11107	10926	10120	10368	9934	10488	9916	6435	6112	5654	-																			
27	*Betacoronavirus*	*Nobecovirus* 1	11151	11331	11233	11268	11414	11332	11360	11252	11267	11340	11255	11344	11328	11140	11154	11349	11252	10453	10624	10418	10645	10023	10178	10234	10344	10182	-																		
28	*Betacoronavirus*	*Nobecovirus* 2	10974	11112	11096	11044	11113	11098	11024	11013	11087	11154	11064	11225	11154	10942	10747	11190	11142	10089	10404	9962	10438	9871	10016	10190	10315	9954	6670	-																	
29	*Betacoronavirus*	*Sarbecovirus* SARS	11087	11161	11143	11138	11300	11210	11210	11130	11222	11277	11310	11345	11282	11069	11038	11401	11205	10419	10616	10440	10757	8874	10006	10001	10192	9983	10077	9944	-																	
30	*Deltacoronavirus*	*Andecovirus*	10687	10679	10784	10780	11045	10768	10695	10811	10774	10837	10872	10959	10914	10957	10512	10848	10770	10922	11116	10725	11157	11105	10930	11025	11123	10829	11135	10924	11197	-																
31	*Deltacoronavirus*	*Buldecovirus* 1	10833	10753	10765	10839	11143	10825	10807	10790	10794	10988	10888	10950	10945	10972	10597	10945	10885	10857	11045	10794	11144	11027	10823	10986	11057	10839	10971	10848	10926	8581	-															
32	*Deltacoronavirus*	*Buldecovirus* 2	10873	10947	10979	10939	11326	10978	11002	11006	10909	11024	11050	10982	11089	11105	10906	11021	10941	11004	11086	11085	11170	11141	11053	11035	11092	11036	11093	11052	11110	8791	5472	-														
33	*Deltacoronavirus*	*Buldecovirus* 3	10854	10962	10936	10976	11316	10981	10945	10940	10852	11067	11068	10973	10973	11107	10826	10973	11042	11048	11107	11025	11203	11202	11020	11058	11054	10993	11079	11004	11067	8712	5574	5063	-													
34	*Deltacoronavirus*	*Buldecovirus* 4	10832	10834	10846	10846	11231	10817	10836	10823	10885	11033	10973	10977	10936	10962	10631	10917	10881	10927	11068	10807	11112	11143	10907	11024	11108	10904	11001	10897	11062	8573	5164	5767	5801	-												
35	*Deltacoronavirus*	*Herdecovirus*	10845	10863	10759	10784	11202	10842	10682	10898	10894	11035	11066	11003	10869	11043	10746	10922	10864	10896	11170	10887	11255	11099	10938	11166	11134	10924	11061	10943	11066	8728	8364	8473	8500	8531	-											
36	*Deltacoronavirus*	*Moordecovirus*	10456	10501	10513	10459	10900	10657	10522	10519	10556	10673	10688	10662	10647	10613	10207	10746	10733	10468	10738	10238	10767	10900	10597	10789	10988	10580	10885	10579	10922	8371	6434	6860	6734	6506	8259	-										
37	*Gammacoronavirus*	*Cegacovirus*	10969	11034	11006	11001	11204	11055	10970	11018	11003	11035	11010	11144	11176	10920	10840	11112	11078	10634	10958	10561	10982	11059	10969	11032	11267	10928	10995	10787	11179	10576	10451	10585	10473	10482	10551	10202	-									
38	*Gammacoronavirus*	*Igacovirus*	10645	10664	10747	10732	11030	10770	10695	10684	10762	10812	10738	10852	10838	10668	10457	10850	10735	10450	10719	10267	10833	10578	10446	10565	10750	10406	10777	10450	10679	10406	10358	10549	10569	10369	10363	9936	9588	-									
39	*Betacoronavirus*	*Sarbecovirus* SARS Cov 2a	10992	10944	10932	11075	11185	11046	11090	10983	11092	11267	11150	11250	11072	10945	10843	11236	11110	10287	10538	10136	10548	8749	9899	9913	10142	9850	10012	9791	4126	11028	10762	11029	10924	10834	10938	10672	11102	10460	-								
40	*Betacoronavirus*	*Sarbecovirus* SARS Cov 2b	10994	10946	10934	11076	11188	11048	11091	10986	11093	11268	11151	11252	11075	10946	10845	11239	11113	10290	10540	10137	10550	8752	9902	9915	10144	9852	10014	9793	4127	11030	10763	11031	10926	10836	10940	10674	11105	10464	0	-							
41	*Betacoronavirus*	*Sarbecovirus* CoV ZC45	11056	11095	11055	11105	11258	11067	11170	11073	11236	11266	11254	11297	11170	10935	10903	11249	11167	10337	10592	10243	10702	8781	9969	9980	10200	9905	10055	9867	3708	11092	10932	11070	11036	10982	11036	10822	11177	10594	2671	2673	-						
42	*Betacoronavirus*	*Sarbecovirus* SARS Cov ZS B	11073	11149	11127	11125	11292	11193	11197	11118	11209	11267	11300	11341	11281	11057	11031	11386	11196	10404	10602	10426	10750	8855	9996	9992	10179	9973	10058	9928	30	11185	10913	11095	11050	11049	11046	10911	11168	10670	4105	4106	3687	-				
43	*Betacoronavirus*	*Sarbecovirus* Bat RaTG13	11005	10980	10954	11060	11211	11044	11065	10994	11113	11298	11143	11284	11086	10949	10859	11236	11120	10304	10558	10165	10607	8769	9916	9915	10190	9857	10058	9830	4176	11028	10801	11035	10957	10846	10915	10716	11112	10486	824	824	2701	4154	-			
44	*Betacoronavirus*	*Sarbecovirus* Pangolin Cov	10990	10978	10995	11021	11144	11020	11045	10954	11136	11232	11209	11225	11092	10919	10865	11268	11054	10274	10502	10157	10534	8745	9852	9914	10116	9858	10026	9801	4291	11008	10772	11041	10909	10876	10891	10674	11057	10472	2056	2057	2898	4265	2066	-		
45	*Betacoronavirus*	*Sarbecovirus* RmYN02	11046	10917	10943	11060	11145	11024	11055	10973	11058	11246	11158	11221	11051	10886	10824	11182	11058	10278	10553	10137	10518	8736	9879	9902	10166	9859	9952	9772	4192	10995	10781	10998	10882	10800	10901	10664	11013	10464	1047	1047	2734	4176	1257	2344	-	

Table S3: Total character differences between the aligned genomes of SARS-Cov-2 and bat coronaviruses RmYN02 and RaTG13

# of taxon	Genus	Subgenus\Name in the Tree		1	2	3
1	*Betacoronavirus*	*Sarbecovirus*	SARS Cov 2a	-		
2	*Betacoronavirus*	*Sarbecovirus*	Bat RaTG13	1138	-	
3	*Betacoronavirus*	*Sarbecovirus*	RmYN02	1803	2032	-

Captions for Tables S1 to S3

Table S1.

Taxonomic sampling of the study and related data.

Table S2.

Total character differences between the aligned genomes of the all coronaviruses from the subfamily Orthocoronavirinae (family Coronaviridae, order Nidovirales) based on the G-block version of the genomic alignment.

Table S3.

Total character differences between the aligned genomes of severe acute respiratory syndrome-related coronavirus 2 (SARS-Cov-2) (MN908947) and bat coronaviruses RmYN02 and RaTG13 (family Coronaviridae, subfamily Orthocoronavirinae, genus *Betacoronavirus*, subgenus *Sarbecovirus*).

Supplemental References

1. S. G. Siddell *et al.*, Binomial nomenclature for virus species: a consultation. *Archives of virology* **165**, 519 (Feb, 2020).
2. S. G. Siddell *et al.*, Correction to: Binomial nomenclature for virus species: a consultation. *Archives of virology* **165**, 1263 (May, 2020).
3. P. J. Walker *et al.*, Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Archives of virology* **164**, 2417 (Sep, 2019).
4. M. J. Adams *et al.*, 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Archives of virology* **162**, 1441 (May, 2017).
5. F. Wu *et al.*, Author Correction: A new coronavirus associated with human respiratory disease in China. *Nature* **580**, E7 (Apr, 2020).
6. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265 (Mar, 2020).
7. R. Lu *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565 (Feb 22, 2020).
8. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270 (Mar, 2020).
9. P. Liu *et al.*, Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathogens* **16**, e1008421 (May, 2020).
10. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nature medicine* **26**, 450 (Apr, 2020).
11. K. Bukhari *et al.*, Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. *Virology* **524**, 160 (Nov, 2018).
12. N. J. Maclachlan, E. J. Dubovi, S. W. Barthold, D. F. Swayne, J. R. Winton, *Fenner's Veterinary Virology: Fifth edition*. (Elsevier Inc, [Place of publication not identified, 2016).
13. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059 (Jul 15, 2002).
14. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772 (Apr, 2013).
15. M. A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1 (2010).
16. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* **17**, 540 (Apr, 2000).
17. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* **27**, 221 (Feb, 2010).
18. I. J. Kitching *et al.*, *Cladistics: The Theory and Practice of Parsimony Analysis*. (Oxford University Press, 1998).
19. E. V. Mavrodiev, C. Dell, L. Schroder, A laid-back trip through the Hennigian Forests. *PeerJ* **5**, e3578 (2017).

20. F.-J. Lapointe, G. Cucumel, The Average Consensus Procedure: Combination of Weighted Trees Containing Identical or Overlapping Sets of Taxa. *Systematic Biology* **46**, 306 (1997).
21. F.-J. Lapointe, C. Levasseur, in *Phylogenetic Supertrees: Combining information to reveal the Tree of Life,* O. R. P. Bininda-Emonds, Ed. (Springer Netherlands, Dordrecht, 2004), pp. 87-105.
22. W. P. Maddison, D. R. Maddison. (Maddison, W. P. & Maddison, D. R., 2018).
23. D. L. Swofford. (Sinauer Associates, Sunderland, MA, 2002).
24. A. Rambaut. (Rambaut, A., 2012).
25. C. J. Creevey, J. O. McInerney, Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* **21**, 390 (2004).
26. C. J. Creevey, J. O. McInerney, Trees from trees: construction of phylogenetic supertrees using clann. *Methods Mol Biol* **537**, 139 (2009).
27. D. M. Williams, M. C. Ebach, *Foundations of Systematics and Biogeography*. (Springer US, 2007).
28. D. M. Williams, M. C. Ebach, *Cladistics: A Guide to Biological Classification*. Systematics Association Special Volume Series (Cambridge University Press, Cambridge, ed. 3, 2020).
29. E. V. Mavrodiev, A. Madorsky, TAXODIUM version 1.0: a simple way to generate uniform and fractionally weighted three-item matrices from various kinds of biological data. *PLoS One* **7**, e48813 (2012).
30. X. Zhou, X. X. Shen, C. T. Hittinger, A. Rokas, Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *Mol Biol Evol* **35**, 486 (Feb 1, 2018).
31. J. Trifinopoulos, L. T. Nguyen, A. von Haeseler, B. Q. Minh, W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research* **44**, W232 (Jul 8, 2016).
32. H. Zhou et al., A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Current Biology* **30**, (2020).
33. L. Pipes, H. Wang, J. Huelsenbeck, R. Nielsen, Assessing uncertainty in the rooting of the SARS-CoV-2 phylogeny. bioRxiv, (2020).