Mathematical Biology and Bioinformatics
2022. V. 17. № 2. P. 289–311. doi: 10.17537/2022.17.289

========== BIOINFORMATICS ==========

On Classification and Taxonomy of Coronaviruses (*Riboviria*, *Nidovirales*, *Coronaviridae*) with Special Focus on Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2)

Mavrodiev E.V.[*1], Tursky M.L.[†2, 3], Mavrodiev N.E.[‡4], Schroder L.[§5], Laktionov A.P.[6], Ebach M.C.[††7], Williams D.M.[‡‡8]**

[1]*University of Florida, Florida Museum of Natural History, Museum Road and Newell Drive, Dickinson Hall, 301, Gainesville, FL, 32611, USA*
[2]*St Vincent's Centre for Applied Medical Research and Department of Hematology and BM Transplant, St Vincent's Hospital, Sydney; NSW, 2010; Australia*
[3]*School of Clinical Medicine, Faculty of Medicine and Health; UNSW Sydney; NSW, 2010; Australia*
[4]*Santa Fe College, Gainesville, Florida, 3000 NW 83 St, Gainesville, FL 32606*
[5]*University of Alaska Anchorage, WWAMI, College of Health, 3211 Providence Drive, HSB 301, Anchorage, AK, 99508, USA*
[6]*Astrakhan State University, Department of Biology, Shaumyan's Square, Building 1, Astrakhan, 414000, Russia*
[7]*Paleontology, Geobiology and Earth Archives Research Centre (PANGEA), School of Biological, Earth and Environmental Sciences, UNSW, Kensington, NSW, 2052 Australia*
[8]*The Natural History Museum, London, SW7 5BD, UK*

Abstract. Coronaviruses are highly virulent and therefore important human and veterinary pathogens worldwide. This study presents the first natural hierarchical classification of *Coronaviridae*. We also demonstrate a "one-step" solution to incorporate the principles of binomial (binary) nomenclature into taxonomy of *Coronaviridae*. We strongly support the complete rejection of the non-taxonomic category "virus" in any future *taxonomic* study in virology. This will aid future recognition of numerous virus species, particularly in the currently monotypic subgenus *Sarbecovirus*. Commenting on the nature of SARS-CoV-2, the authors emphasize that no member of the *Sarbecovirus* clade is an ancestor of this virus, and humans are the only natural known host.

Key words: *Coronaviridae, cladistics, rooted phylogenetic trees, taxonomy, binominal nomenclature, generic circumscription, SARS-CoV-2.*

INTRODUCTION

Coronaviridae is a virus family of the order *Nidovirales* (realm *Riboviria*) [1, 2]. According to the current summaries of the International Committee on Taxonomy of Viruses (ICTV), this family is divided into two subfamilies – *Letovirinae* and *Orthocoronavirinae* [1–

[*] Address for correspondence: evgeny@ufl.edu
[†] m.tursky@amr.org.au
[‡] nickmavrick11@gmail.com
[§] lamille3@gmail.com
[**] alaktionov@list.ru
[††] mcebach@gmail.com
[‡‡] DavidMyWilliams@gmail.com

7]. These two subfamilies circumscribe five genera: the monotypic genus *Alphaletovirus* of the subfamily *Letovirinae* (with the single species *Microhyla letovirus* 1), and four non-monotypic genera of the subfamily *Orthocoronavirinae*, namely:

1. genus *Alphacoronavirus* with 12 subgenera: *Colacovirus*, *Decacovirus*, *Duvinacovirus*, *Luchacovirus*, *Minacovirus*, *Minunacovirus*, *Myotacovirus*, *Nyctacovirus*, *Pedacovirus*, *Rhinacovirus*, *Setracovirus*, and *Tegacovirus*;
2. genus *Betacoronavirus* with five subgenera: *Embecovirus, Hibecovirus, Merbecovirus, Nobecovirus,* and *Sarbecovirus*;
3. genus *Gammacoronavirus* with two subgenera: *Cegacovirus* and *Igacovirus*;
4. genus *Deltacoronavirus* with four subgenera *Andecovirus, Herdecovirus, Moordecovirus* and *Buldecovirus* (Table S1)[1–7].

The viruses of *Coronaviridae*, such as the severe acute respiratory syndrome (SARS-CoV) and related Middle East respiratory syndrome (MERS-CoV), are highly virulent [8]. 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. About two years and seven months later, the total cases of COVID-19 was estimated to be 632,533,408; and cumulative deaths have exceeded 6,592,320 (www.who.int, accessed November 15, 2022). Thus, *Coronaviridae* is of considerable medical importance worldwide. Such importance has resulted in some urgency to further understand the relationships within the coronavirus family, and the viruses most closely related to SARS-CoV-2.

On the motivations, goals, semantic frames and the novelty of the study

In conjunction with the current coronavirus pandemic, interest towards virological literature has grown, especially among biologists from various non-virological fields. These include the authors of this paper. However, we, as well as many other scientists, often find published coronavirus-related texts, especially phylogenetic trees, almost unreadable and difficult to interpret. Thus, the objective of this paper is to help virologists effectively and efficiently resolve these issues. In other words, this study serves as feedback from a different part of the scientific community during the critical time of the global pandemic.

We believe the main reasons for difficulties in both virologic texts and phylogenetic analyses to be:

1. Continually employing phylogenetic trees that are based on poor taxonomic sampling, which also typically lack basal rooting.
2. Constant inconsistent usage of cumbersome abbreviations and trivial names of different coronaviruses by virologists (including members of the International Committee on Taxonomy of Viruses (ICTV) and *Coronaviridae* Study Group (CSG)) in various publications.
3. Wide acceptance of the non-taxonomic category “virus” in different taxonomic studies in the field of virology.
4. Attempts to incorporate the genomes of numerous viruses (and coronaviruses in particular) in analyses without having the proper taxonomic and phylogenetic framework, as well as having a natural classification of the family *Coronaviridae*.
5. A strict focus on a single phylogenetic method – specifically, the Maximum Likelihood (ML) approach.

Thus, in this study we argue that:

1. Only basally rooted and well-taxonomic sampled trees should be used in any analyses of the *Coronaviridae*.
2. We agree with authors who have already suggested that the creation of binomial nomenclature for viruses should not be difficult [9]. In particular, we demonstrate a “one-step” solution to incorporate the principles of binomial nomenclature into the taxonomy of *Coronaviridae*.

3. We strongly support the rejection of the non-taxonomic category “virus” in any future *taxonomic* study of *Coronaviridae*. This will aid in recognition of numerous undiscovered virus species, particularly, in the currently monotypic subgenus *Sarbecovirus*.

4. The first Natural Classification of *Coronaviridae* has been established in this study (Figures 1–3). As described below, numerous genomes of different viruses from any clade of *Coronaviridae* can be used in separate clade-based phylogenetic analyses of the family. The obtained trees can be combined through the use of various supertree methods.

5. In addition to the parametric (model-based) ML method, we have used the following three 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 pattern-cladistic (or Cladistic) approaches have not previously been used to resolve the relationships within *Coronaviridae*, or within any other virus family. However, because much of the discussion within the field of taxonomy of viruses is actually focused on classification [1–7], methods of Cladistic analysis may be advantageous to virologists.

We would like to stress that this study is not a standard taxonomic work in the field of virology. We have deliberately proposed no nomenclature changes within *Coronaviridae*. For several reasons (see above), we decided to establish an alternate perspective on this topic that may be beneficial to experts in virology. Once again, our paper does not describe new tribes, genera, or species. Rather, it is intended to present efficient and practical ways of doing so to virologists-taxonomists. The key ICTV and CSG solutions within *Coronaviridae* are summarized in our Table S1 and Figures 1–3 as well as throughout the paper. Based on this, we believe our paper fits the current ICTV taxonomic standards.

The novelty of our work is a cladistic or purely comparative view of the nature of the recently discovered coronavirus SARS- CoV-2. We propose a completely novel statement of the endemism of SARS-CoV-2, as well as pointing out gross aggregated speculations regarding the bat or pangolin origins of this coronavirus, as fairly questionable. It is worth stressing that no member of the *Sarbecovirus* clade is an ancestor of SARS-CoV-2, and the notion of transition of this virus from animals such as bats, or pangolins to humans, is problematic.

The sampling of phylogenetic studies of *Coronaviridae* and rooting of the obtained trees

Coronaviridae is a virus family with a potentially high amount of undiscovered diversity at the species level [3, 10].

Based on multiple comprehensive phylogenetic analyses of the ICTV-approved genomes of 39 different species of coronaviruses, we have tested the monophyly and relationships of all current subgenera of *Coronaviridae*, as well as critical proposals of the interrelationships of the subfamilies.

Many previous phylogenetic analyses of *Coronaviridae* have included the use of unrooted or mid-point rooted trees, some include viruses from other families, and in many cases use a very limited taxonomic sampling [3, 11–14].

Basal rooting of cladograms is critical to resolving phylogenetic relationships accurately [15–17]. Within conventional phylogenetic frameworks, the root of the cladogram can be defined as the most basal taxon of the cladogram on which all characters were polarized [15]. Thus, rooting of a cladogram can be performed either *a priori* or *a posteriori* to the results of the analyses [15]. Basal rooting is a means of restructuring data to yield a more stable and rigorous hierarchical classification. By comparison, procedures such as mid-point rooting rely on the artificial assumptions of the molecular clock and are unnecessary if the outgroup taxon of some particular relationship is actually known or accurately assumed [15]. The simple presence of other viruses in the working matrices of *Coronaviridae* is not an assurance of the basal rooting of the resultant trees, and there is no guarantee that the trees will appear to be

rooted "automatically" if no distant relatives of coronaviruses have been included in the analyses [12, 13]. For example, in the application of phylogenetic methods to the molecular matrices of *Coronaviridae* and prospective relatives, the resultant tree(s) must still be rooted relative to one of the *a posteriori* selected outgroups. By operating this way, the shape of the final tree may be dramatically changed as compared to the initial outputs of the analysis. For *Coronaviridae,* suitable outgroup taxa are known and can be used for basal rooting, resulting in rooted trees that will represent a more stable hierarchical classification of the family.

On the cladistic analyses of *Coronaviridae*

Almost all phylogenetic studies of coronaviruses so far have been based on the parametric Maximum Likelihood method. In order to increase the veracity of representations of the relationships within *Coronaviridae*, in addition to the ML method, we have used the three non-parametric cladistic approaches mentioned above.

The following provides some formal definitions to explain the relevance of the methods chosen in this study. "Phylogenetic tree" is a hypothesis of *genealogical relationships* among a group of taxa with specific connotations of common ancestry and a time axis [15]. Most trees that have been obtained in the phylogenetic studies of *Coronaviridae* are defined as "phylogenetic trees" [3, 11–14, 18].

However, a more general term is a "cladogram", which is a branching diagram specifying hierarchical relationship among taxa based upon homologies [15–17]. For simplicity, a cladogram can also be called a "tree" [15]. However, contrary to a phylogenetic tree *per se*, a cladogram includes no references to common ancestry and implies no time axis [15–17].

Within conventional cladistics analysis [15, 19], the cladogram may also be treated as a phylogenetic tree [19]. However, this is not necessary within the more general interpretation of cladistics [16, 17, 20–23] that is flexible for use in process-based explanations of the observed taxonomic patterns and its hierarchies. Maximum Parsimony may also be considered a method of classification that groups taxa hierarchically into nested sets, representing sets such as cladograms [15, 20, 22].

The major goal of cladistic analysis is the search for a natural hierarchy of patterns: clades or monophyletic groups (relationships that are based solely on synapomorphies). Essentially, cladistics is just an extension of the comparative thinking that forms the heart of traditional biology (such as zoology or botany) [16, 17, 20–23].

By comparison, as was clearly stated by Felsenstein, the ML estimation is an evolutionary model-based method of statistical inference that involves finding the phylogenetic tree that yields the highest probability of "evolving" the observed data [24, 25]. All data (characters), not solely synapomorphies (or homologies), are therefore used in ML analyses (and other phenetic procedures) [16, 26].

Monophyly and major hypothesis on hierarchical relationships within *Coronaviridae*

The trees produced from these four methods enable testing of the monophyly of all genera and infrageneric taxa of *Coronaviridae* and determine whether they are representative of a natural hierarchy.

The first assessment of monophyly was that of the newly described monotypic genus *Alphaletovirus* (Table S1), a member of subfamily *Letovirinae*, as a sister group of the family *Coronaviridae* [11].

The second monophyly assessment is the general relationship within the subfamily *Orthocoronavirinae* as the simple hierarchy:

(((((*Alphacoronavirus*),(*Betacoronavirus*))(*Gammacoronavirus*))(*Deltacoronavirus*))) [27].

This relationship is named the ((((A,B)Γ)Δ)) hypothesis of the general relationship within the subfamily.

Thirdly, the subgenus *Hibecovirus* has been placed as a sister taxon of the presumably monotypic subgenus *Sarbecovirus* [3, 28]. This result is based on unrooted phylogenies, so here it has been tested with special attention to the placement of SARS-CoV-2.

Fourthly, the close relationship of SARS-CoV-2 with bat coronaviruses RaTG13 and RmYN02, as well as with pangolin coronavirus, was tested [29–32].

Discussion points regarding virology nomenclature

The current nomenclature for the family *Coronaviridae* was developed by the CSG and approved by the ICTV (see above). Around 2010 the CSG paused introduction of binomial nomenclatures for viral species to allow discussion. Revision of the nomenclature for viral species is likely to resume. It is therefore pertinent to consider how the principles of binomial nomenclature could be applied to *Coronaviridae*, and how this would impact future recognition of virus species, in particular those in the current monotypic subgenus *Sarbecovirus*.

We strongly support the adoption of binomial nomenclature for viruses [6, 7, 9]. The binomial nomenclature would be relatively easy to implement and would be clear and practical to understand by people from a variety of fields [6, 7, 9]. We believe this would reduce difficulties in the interpretation and synthesis of results arising from the usage of cumbersome abbreviations and trivial names of different coronaviruses used by virologists in various scientific publications, past and present.

We also suggest the rejection of the non-taxonomic category “virus” in any future taxonomic research in virology, even if in this study we use this term following the current rules of ICTV.

Our suggestions herein do not seek to create conflict but to offer options for discussion to minimize conflict while providing clarity in the future. For example, we offer for discussion an introduction of the rank of tribe. Our paper does not formally describe new tribes, genera, or species, but aims to demonstrate the potential efficient and practical binomial virus nomenclature for consideration in the determination of future nomenclature decisions.

MATERIALS AND METHODS

Taxonomic sampling of the study

39 ICTV-approved genomes of species of *Coronaviridae* have been used in this study, with essential information including virus names, corresponding abbreviations, suggested hosts (if any), GenBank numbers, etc., summarized in Table S1 as was previously done in [33].

In addition, the following were included with the final alignments:

1. The published genome MN908947 of virus SARS-CoV-2 [14].
2. An unpublished genome of the same virus species (MN988713).
3. The genome of bat SARS-like coronavirus (bat-SL-CoV-ZC45; MG772933) previously used in other published comparisons to SARS-CoV-2 [13, 14].
4. Bat coronavirus RaTG13 (MN996532), pangolin coronavirus (MT121216) and RmYN02 (GISAID: EPI_ISL_412977), previously established to be the closest known relatives of SARS-CoV-2 [29–32].
5. An additional SARS related unpublished genome ZS-B (AY394996).
6. The genome of *Microhyla letovirus* 1 or MLeV virus (subgenus *Milecovirus*), obtained upon courtesy request from Prof. B. W. Neuman (Texas A&M University-Texarkana, TX, US).
7. The ICTV-approved genome of *Torovirus* (AY427798) (family *Tobaniviridae*, order *Nidovirales*)[34]. The later was assumed to be the best outgroup taxon of *Coronaviridae* [8].

The names of major relationships of *Coronaviridae* obtained from the trees produced in this study are frequently written in italics due to the strong congruence with different

taxonomic entities. Depending on the context, the names of the clades are provided in both square 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.

Alignments

All genomic alignments have been performed using MAFFT following FFT-NS-I strategy with the command: mafft --inputorder --adjustdirection --anysymbol -- kimura 1 --maxiterate 1000 --6merpair input [35–37].

To remove poorly aligned positions from the obtained genomic alignments, we used the simple, easy and efficient program G-block [38] as implemented in SeaView [39], thereby solving the saturation problem with the alignment of virus genomes [40] directly by exclusion of saturated sites from the molecular matrix. Moving these sites out of alignment makes future analyses relatively noise-less [38]. Based on the high accuracy and efficiency of this method, we focused on the G-Block-based whole genome alignment of *Coronaviridae* + *Torovirus*, rather than be limited to the codon-based alignment of the same group of taxa.

Under the conditions of the "less stringent" strategy of the G-block algorithm [38, 39], the saturated positions have been successfully removed from

1. the genomic alignment of *Coronaviridae* + *Torovirus*, and
2. the genomic alignment of subfamily *Orthocoronavirinae* with no outgroups included [38, 39].

The G-block version of the genomic alignment of *Coronaviridae* + *Torovirus* was also established as a binary matrix using simple "presence – absence" coding with the future inclusion of the "all-plesiomorphic" ("all-zero") artificial taxon. In short, G-block based genomic alignment of the family *Coronaviridae* + *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), this can be written as "1000", where "1" means "the A is present in position 253", and subsequent "0" indicates that U(T), G and C are simultaneously absent on the same position [15, 21].

Assuming 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-zero" outgroup added was later used as an input into the script Forester v. 1.0 following the command ruby trees.rb path_to_matrix_file with future selection of the "Additional" forest of the maximal trees (relationships) for Average Consensus (AC) analysis [21, 41, 42].

Manipulations of matrices and the tree-files have been performed with Mesquite v. 3.70 [43], PAUP* v. 4.0a 169 [44] and FigTree v. 1.4.4 [45].

Trees and analyses

The G-block version of the molecular alignment of *Coronaviridae* + *Torovirus* was analyzed using MP (Fitch Parsimony) [reviewed in 15], and by 3TA with fractional weighting [15–17, 23, 46, 47].

Following the logic of Williams-Siebert (WS) representation of the unordered multistate data, 3TS permutations of the G-block-based alignment of *Coronaviridae* + *Torovirus* were conducted using TAXODIUM version 1 with the command: taxodium.exe input_file_name.csv –idna –ob –og –fw –nex and taking values of the operational outgroup as equal to the values of *Torovirus* [47, 48].

All MP analyses were performed using PAUP*; the resulting most parsimonious tree was *a posteriori* rooted relative to *Torovirus* [21, 44, 47].

The AC of the array of maximal trees was calculated using the program Clann version 4.1.5 as suggested in previous studies [21, 49, 50]. The distance optimality criterion for the AC analysis was specified in the simplest way as a "distance with non-weighted least squares" [41, 42, 44, 49].

Following others [18, 51], ML analysis of G-block alignment of *Coronaviridae* + *Torovirus* was conducted with W-IQ-TREE [52] with implemented automatic model selection procedure. The resulting most probable tree was *a posteriori* rooted relative to *Torovirus*.

The MP Bootstrap support (BS) values have been calculated as described earlier [47]. In the case of the ML analysis, the Approximate Likelihood ratio Tests (aLRT) [53] support values have been calculated instead of the ML BS supports, as implemented W-IQ-TREE.

The simplest "total" character difference (TCD) [25, 44] between the G-block modified aligned genome sequences of subfamily *Orthocoronavirinae*, 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* [44] under the default options. The TCD 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 [44]. 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.

RESULTS

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

The genomic alignment of subfamily *Orthocoronavirinae* with no outgroup included is 49,881 bp. The G-block version of this alignment consists of 23,431 bp. This alignment was used to calculate TCD between all 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 TCD between newly discovered SARS-CoV-2 and two of its closest relatives (Table S3).

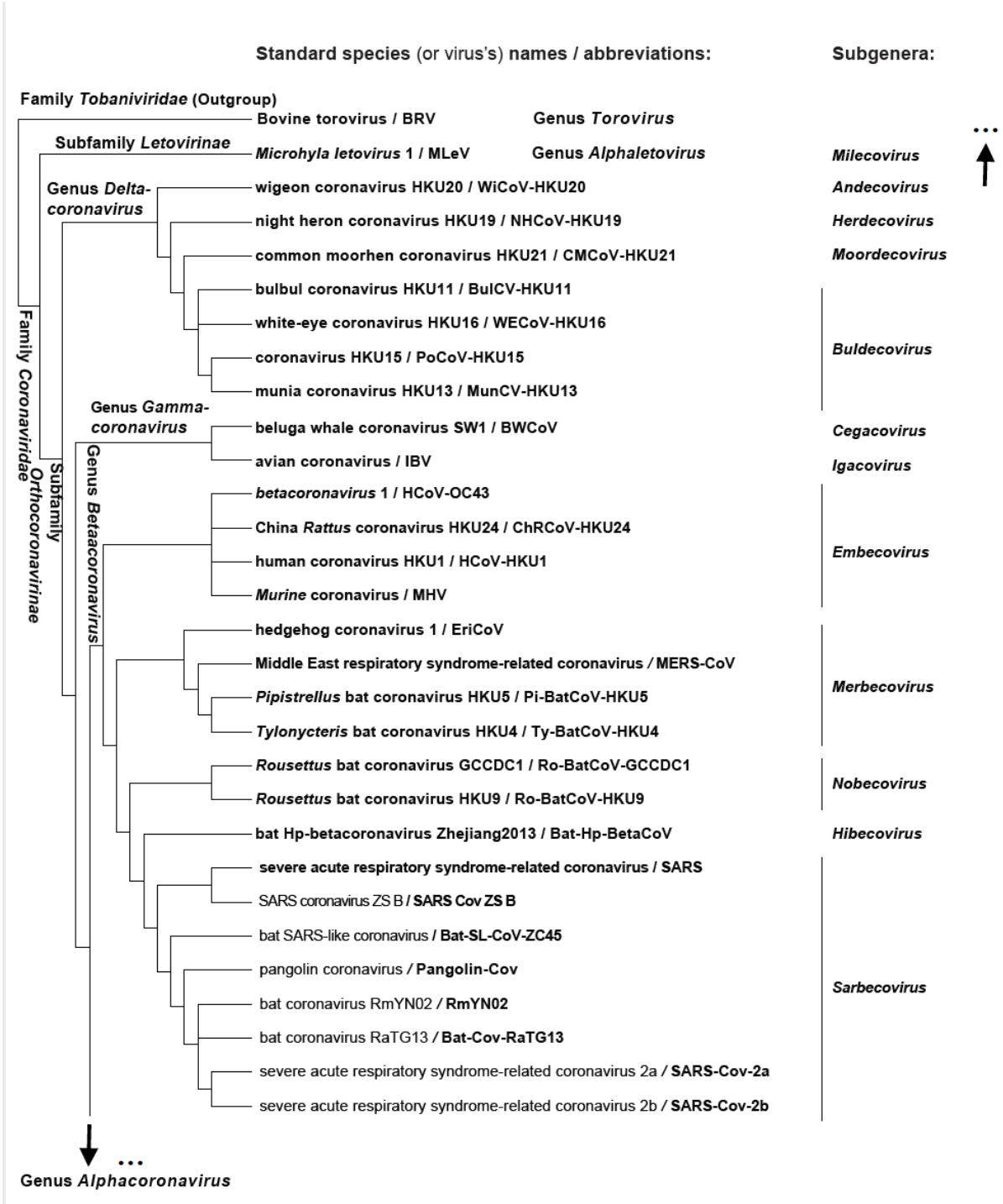
All four trees have several major consistencies, as shown in the Strict Consensus (Figures 1–3, S1–S4).

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 3TA representation of the same 22,489 bp G-block alignment resulted in 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).

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

All obtained alignments contain a huge number of variable characters and every *Coronaviridae* virus seems to be separate from the others sometimes by hundreds or (more commonly) thousands of SNPs (Tables S2 and S3). For example, the minimal relationship

{RmYN02 {RaTG13 + SARS-CoV-2}}, based on the 29,907 bp complete genomic alignment of three taxa (RmYN02, RaTG13 and SARS-CoV-2), implies 1,329 informative SNPs (or, respectively, 1,329 3TSs) from the total 2,467 variable characters.



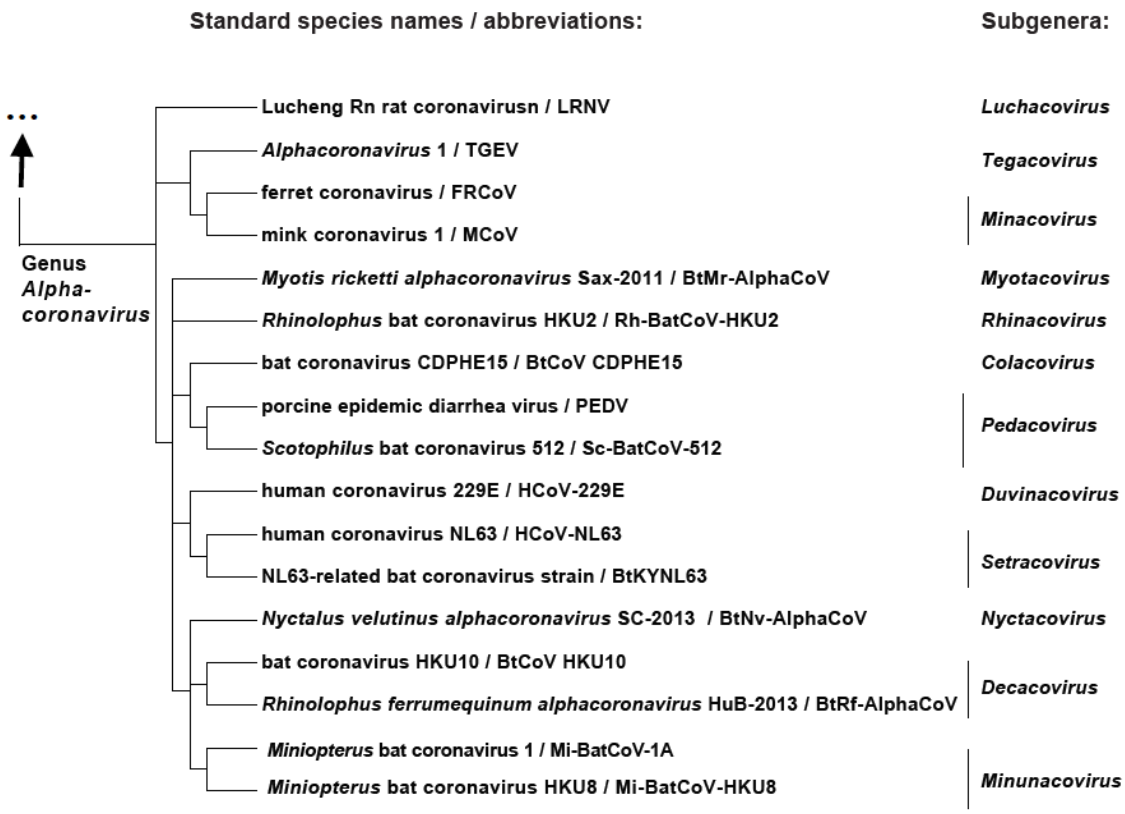


Fig. 1. Strict Consensus of four trees produced by three different methods of cladistic analysis as well as by the Maximum Likelihood method using genomic alignment of *Coronaviridae* + *Torovirus* drawn following standard virus names and abbreviations. Saturated sites have been removed from the genomic alignment before analyses. Standard virus names (not in bold) have been used only for a few viruses from subgenus *Sarbecovirus*. See Figures S1–S4 for more detail including the tree node support values.

General patterns of relationships within *Coronaviridae* and the placement of *Microhyla letovirus* 1

The results of all analyses 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 the family.

Hierarchical classification of *Coronaviridae* established as a simplified Strict Consensus of four trees (Figure 3) was produced by three different methods of cladistic analysis as well as by ML approach.

Below, the individual clades and major relationships within the *Coronaviridae* are described.

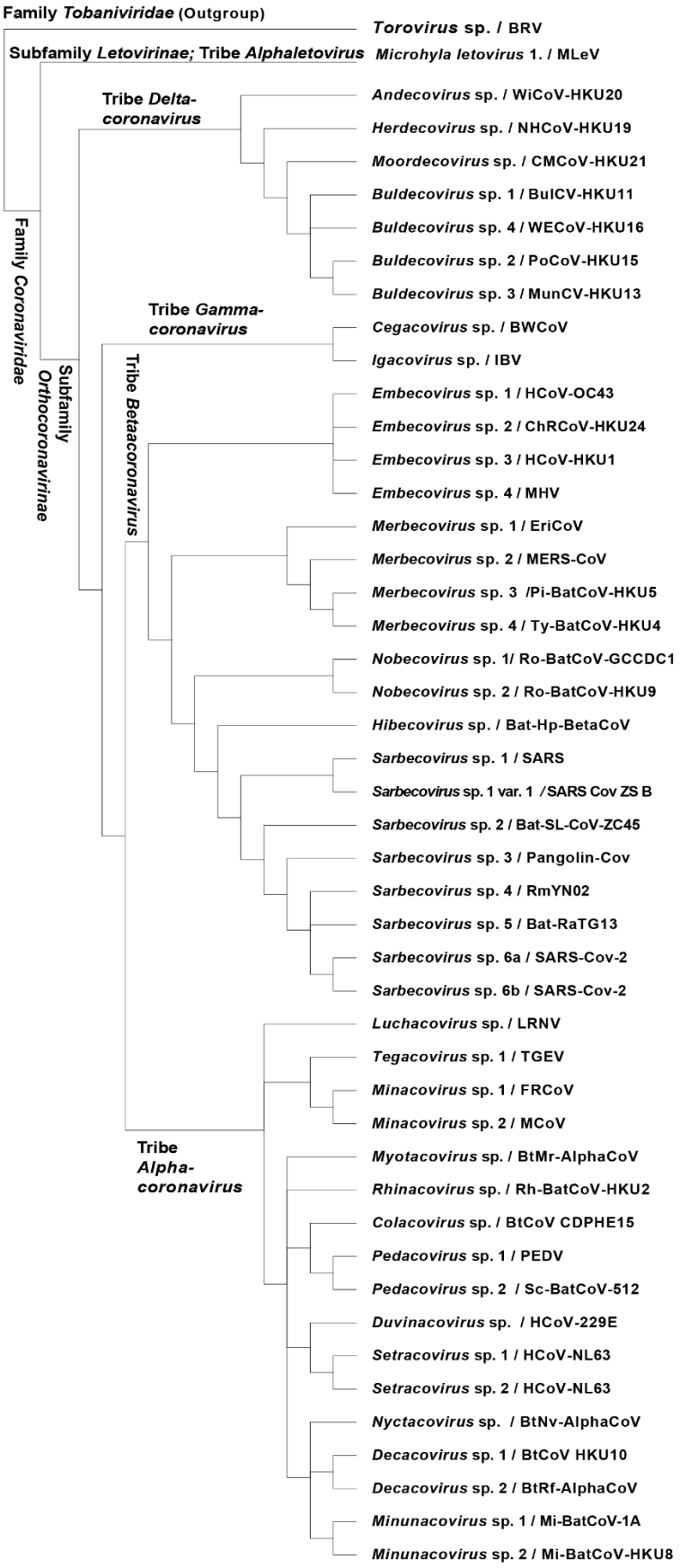


Fig. 2. The same consensus tree as Figure 1, with the suggested changes in the nomenclature of family *Coronaviridae*. For simplicity, the names of prospective tribes have been spelled according to the names of the current genera (e. g. tribe *Betacoronavirus*). The names of the viruses such as SARS-CoV are given as binomials (e. g., *Sarbecovirus* sp. 1).

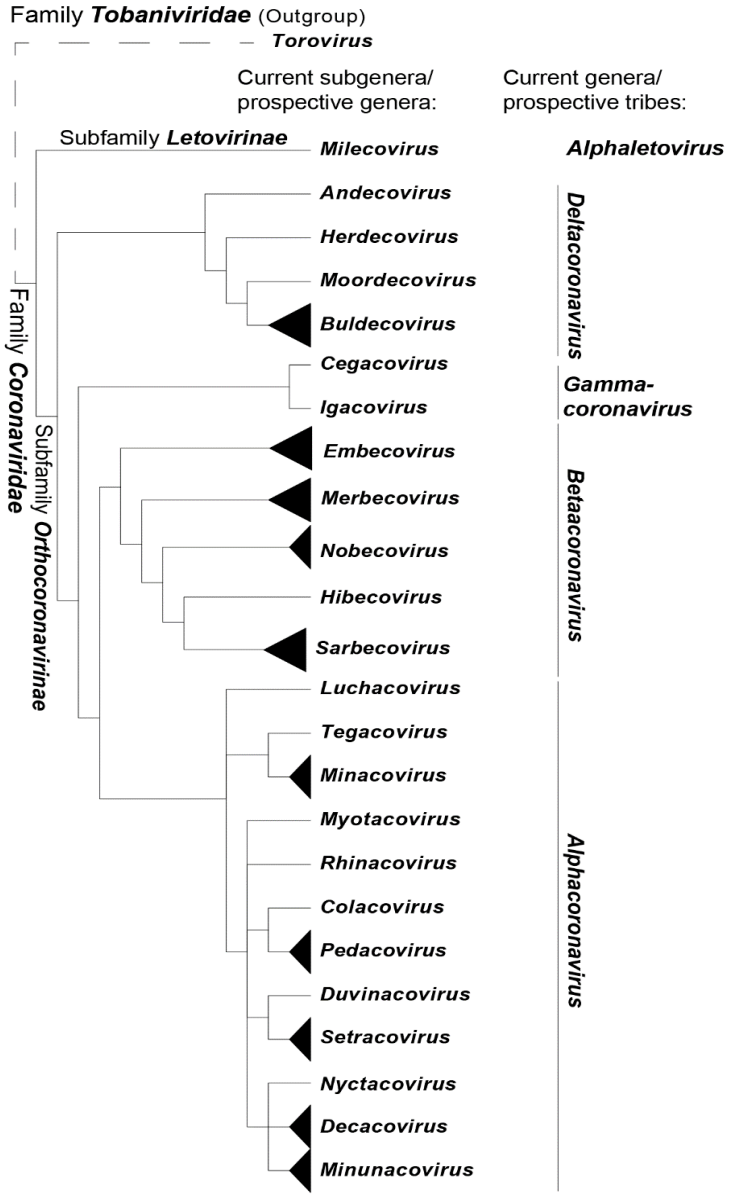


Fig. 3. Hierarchical classification of the coronaviruses (*Riboviria*, *Nidovirale*s, *Coronaviridae*) given as simplified Strict Consensus of four trees produced by three different methods of cladistic analysis as well as by Maximum Likelihood method using modified genomic alignment of *Coronaviridae* + *Torovirus* (see Figure 1 for more details).

{*Alphacoronavirus*}

Within the clade {*Alphacoronavirus*}, all four trees show the following taxa as sisters:

1. ferret coronavirus and mink coronavirus 1 (both taxa are from subgenus *Minacovirus*);
2. porcine epidemic diarrhea virus and *Scotophilus* bat coronavirus 512 (both taxa are from subgenus *Pedacovirus*);
3. human coronavirus NL63 and NL63-related bat coronavirus BtKYNL63-9b (both taxa are from subgenus *Setracovirus*);
4. bat coronavirus HKU10 and *Rhinolophus ferrumequinum* alphacoronavirus HuB-2013 (both taxa are from subgenus *Decacovirus*);
5. *Miniopterus* bat coronavirus 1 and *Miniopterus* bat coronavirus HKU8 (subgenus *Minunacovirus*).

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 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).

Thus, within {*Alphacoronavirus*}, we were able to find five smaller clades (subclades): {*Decacovirus*}, {*Minacovirus*}, {*Minunacovirus*}, {*Pedacovirus*} and {*Setracovirus*} exactly corresponding to each of these clades with the previously established subgenera of *Alphacoronavirus*.

From our results the following is clear:

1. subgenus *Colacovirus* (bat coronavirus CDPHE15) is a sister to *Pedacovirus* clade in all of the trees;
2. subgenus *Duvinacovirus* (human coronavirus 229E) is a sister to the *Setracovirus* clade in all of the trees; and
3. subgenus *Tegacovirus* (alphacoronavirus 1) is a sister to 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. It should be strongly emphasized, however, that standard MP, as well as Cladistic methods, but not the ML method, have placed *Luchacovirus* as a sister of the {*Alphacoronavirus*}.

{*Betacoronavirus*}

All of the analyses argue in favor of the general simple hierarchical relationship within {*Betacoronavirus*}:

({*Embecovirus*}({*Merbecovirus*}({*Nobecovirus*}(*Hibecovirus*,{*Sarbecovirus*})))).

The relationships of four species of the subgenus *Embecovirus* (betacoronavirus 1, China rattus coronavirus HKU24, human coronavirus HKU1 and murine coronavirus), that formed a clade with the same name, the sister of the clade, that contains all the remaining subgenera of *Betacoronavirus*, depends on the method of the analysis and appeared as unresolved on the Strict Consensus.

All four species of subgenus *Merbecovirus* form a same-name clade. In all trees, the hedgehog coronavirus 1 is a sister to the clade that contains three other members of the subgenus *Merbecovirus*: namely, Middle East respiratory syndrome-related coronavirus, *Pipistrellus* bat coronavirus HKU5, and *Tylonycteris* bat coronavirus HKU4.

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

Clade {*Sarbecovirus*-SARS + SARS Cov ZS B} 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 coronavirus SARS-CoV-2. The pangolin coronavirus (isolate MP789) has been defined as a sister of the clade {RmYN02 + RaTG13 + SARS-CoV-2} in all of the analyses.

All trees define the monotypic subgenus *Hibecovirus* as a sister of {*Sarbecovirus*}.

Two members of subgenus *Nobecovirus*, namely Rousettus bat coronavirus GCCDC1 and *Rousettus* bat coronavirus HKU9, are sister taxa.

{*Gammacoronavirus*}

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 in all of the analyses.

{*Deltacoronavis*}

Five subgenera of genus *Deltacoronavis*, namely subgenus *Andecovirus* (with single species wigeon coronavirus HKU20), subgenus *Buldecovirus* (with four species: bulbul coronavirus HKU11, coronavirus HKU15, munia coronavirus HKU13 and white-eye coronavirus HKU16), monotypic subgenus *Herdecovirus* (Night heron coronavirus HKU19) and monotypic subgenus *Moordecovirus* (with common moorhen coronavirus HKU21), have formed the {*Deltacoronavis*} clade in all of the analyses. Also, all of the analyses argue in favor of the simplest hierarchy of the relationships within this clade:

(*Andecovirus* (*Herdecovirus* (*Moordecovirus* ({*Buldecovirus*})))).

Within the clade {*Buldecovirus*}, coronavirus HKU15 and munia coronavirus HKU13 appeared to be sisters in all of the analyses; the relationship between the other members of the {*Buldecovirus*} depends on the method of the analysis. Monotypic subgenus *Andecovirus* (wigeon coronavirus HKU20) is a sister of {*Deltacoronavis*}.

Monophyly of non-monotypic taxa of *Coronaviridae*

As demonstrated above, all four trees show that all four current genera of subfamily *Orthocoronavirinae*, namely *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus*, and *Gammacoronavirus*, are monophyletic (Figures 1–3, S1–S4). All non-monotypic subgenera of all four genera of *Orthocoronavirinae* are monophyletic in all of the analyses.

DISCUSSION

On the general relationships within *Coronaviridae*

The basal rooted trees produced by the four methods herein allow for the effective testing of various suggestions regarding the relationships of viruses within the *Coronaviridae*.

1. This confirms that two subfamilies of *Coronaviridae*, namely subfamily *Letovirinae* (family *Abyssoviridae*) and subfamily *Orthocoronavirinae*, are sisters [11]. This solution, which was previously based on the unrooted trees and limited taxon sampling, is consistent with the familial rank of both taxa; we would recommend accepting the monotypic subfamily *Letovirinae* at the familiar rank [11].

2. Our results clearly argue in favor of the ((((A,B)Γ)Δ)) hypothesis of the general pattern of relationship within subfamily *Orthocoronavirinae* [27]. With basal placement of subgenus *Milecovirus* (MLeV), we can extend this pattern to the relationship: ((((A,B)Γ)Δ),MLeV)). This natural hierarchy within family *Coronaviridae* is in principle congruent to the general pattern of their hosts: (((Mammals) Birds + Mammals) Amphibia)) (Table S1).

3. Subgenus *Hibecovirus* has been treated as a sister taxon of the subgenus *Sarbecovirus* [28]. All four analyses (Figure S1–S4) validate the sister relationship of the subgenera *Hibecovirus* and {*Sarbecovirus*}.

4. Due to particular interest in the virus SARS-CoV-2 across diverse scientific fields, comment on the relationship to SARS-CoV-2 are of interest. Contrary to some recent suggestions, all four methods of phylogenetic analysis place pangolin coronavirus as a sister to the relationship {RaTG13 + RmYN02 + SARS-CoV-2}(Figure S1–S4), confirming that bat coronaviruses RaTG13 or RmYN02, not pangolin coronavirus, are indeed the closest relatives of SARS-CoV-2 [29-32].

Differences in details within clade relationships do exist between the ML method and each of the cladistic methods newly applied to molecular sequence data of *Coronaviridae*. For example, the MP, AC and 3TA trees, but not ML trees, 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 weekly supported sister of the SARS-CoV-2. However, both cladistic methods (3TA and AC, as applied to the array of the basal rooted trees, each corresponding to the binary representation of the standard molecular characters), in contrast, placed bat coronavirus RaTG13, but not RmYN02--as a sister of this newly discovered coronavirus.

All four phylogenetic analyses are initially based on a common molecular matrix (22,489 bp G-Block version of the 47-genomes alignment). Differences between analyses are likely to be the result of how each method deals with conflict when forming optimal trees. Nevertheless, the similarity between 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.

The obtained trees can be easily used for the future analyses of the massive molecular sequence data to clarify the relationships within *Coronaviridae* in greater detail. For example, as described in this paper, numerous genomes of different viruses from {*Sarbecovirus*} (or any other clade, such as {*Merbecovirus*} or {*Decacovirus*}) can be used in separate future phylogenetic studies, including the sister groups, clearly specified in this paper.

Utilizing the trees resulting from our current analyses, all potentially obtainable basally rooted clade-based trees (e.g., the future multi-genomes phylogeny of {*Sarbecovirus*} with *Hibecovirus* specified as an outgroup, the future multi-genomes phylogeny of {*Merbecovirus*} with one of the genomes of the viruses from {*Embecovirus*} clade selected as an outgroup, etc.) can then be simply combined with each other in the following various available supertree methods.

On origin and endemism of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2)

From the elementary comparative (or static) point of view, every currently recognized *Coronaviridae* species seems to be well defined, with each separated from the others by thousands of SNPs (Tables S2 and S3). The closely related viruses from the *Sarbecovirus* clade (including the newly discovered SARS-CoV-2, bat coronaviruses RaTG13 and RmYN02), are all remarkably different from one another from a comparative molecular standpoint [56] (Table S3). The same is true for every relationship within *Coronaviridae* recovered in these analyses.

Such simple observations automatically exclude the possibility of a recombination-based origin of SARS-CoV-2, as well as other semantically similar propositions, such as the laboratory engineering theory of the origin of this virus [57, 58].

Keeping in mind that the proximal origin of SARS-CoV-2 remains unknown [55, 56], and based on the data available to date [55], including the comprehensive trees produced herein (Figures S2, S3), the focus should shift to the static aspect of the problem.

First it is worth stressing that arranging sister taxa as a sequence of ancestors and descendants goes against a cladistic way of thinking [16, 17, 21–23]. In conventional cladistics, an "ancestor" is a purely hypothetical entity, associated with the nodes of the cladogram, not any taxon included in the analysis [21]. Within the framework of pattern-cladism, assuming ancestral taxa as nodes of a cladogram is simply unnecessary [16, 17, 21–23]. Thus, from the cladistic viewpoint, no member of any clade of *Coronaviridae* is an ancestor of any given virus species. For example, either bat coronavirus RmYN02 or RaTG13 can be treated as *sister* of SARS-Cov-2, but not as its *ancestor*. Following this logic, no member of the *Sarbecovirus* clade can be considered an ancestor of SARS-CoV-2.

Next, the monophyly of all genera of *Coronaviridae*, of its non-monotypic subgenera, as well as the general relationship ((((A,B)Γ)Δ)) within the subfamily *Orthocoronavirinaeis,* can also be accurately demonstrated within a clear static analytical framework (Figures S2, S3).

This comparative view may be critical in discussing the general simple pattern in the hosting of SARS-CoV-2.

Recent studies of SARS-related coronaviruses have suggested that bats harbor close relatives to SARS-CoV-2, and that pangolins may be natural hosts of this member of *Betacoronavirus*, leading to the hypothesis of animal-to-human transmission of SARS-CoV-2 [13, 32, 55, 59, 60]. However, the search for other hosts, as well as related exotic ways of transmission of SARS-CoV-2 from these hypothetical hosts to humans, is based on a set of 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..." [60]) and ignores the simple possibility of *original* human-based natural hosting of SARS-CoV-2. The last proposal is semantically similar to the propositions of original bat-based natural hosting of the severe acute respiratory syndrome-related coronavirus (SARS-CoV), or its closest relatives (Table S1).

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 as the latter virus has not yet been detected in animals such as bats or pangolins [59]. As was demonstrated, the spike glycoprotein of SARS-CoV-2 has a broad host tropism for mammalian ACE2 receptors, but bat and bird ACE2 proteins were the least efficiently used receptors, if compared to ACE2 proteins of dogs, cats and other mammals [61].

The possibility of original human hosting of SARS-CoV-2, or, in other words, the endemism of recently discovered SARS-CoV-2, unfortunately has yet to be discussed in scientific literature. Furthermore, the extremely high contagiousness of SARS-CoV-2 is arguably unlikely to have arisen within ca. two years of transmission from animals to humans. The alternative is that a form of SARS-CoV-2 may have existed in some parts of the human population before the beginning of the pandemic of 2020 – 2022 (either within a geographically isolated group or a particular age group, such as infants) but was less pathological or, perhaps, more effectively suppressed by the human immune system for some time prior to the acquisition of increased virulency.

Seven human coronaviruses have been identified to date [3, 62–68]. Examples of four common, human endemic coronavirus are:

1. human coronavirus 229E from subgenus *Duvinacovirus* (genus *Alphacoronavirus*),
2. human coronaviruses OC43 (HCoV-OC43) from subgenus *Embecovirus* (genus *Betacoronavirus*),
3. human coronavirus HKU1 (HCoV-HKU1) from the same subgenus, and
4. human coronavirus NL63 (HCoV-NL63) from subgenus *Setracovirus* (genus *Betacoronavirus*) (Table S1).

These are globally distributed viruses where no animal hosts have ever been proposed that are frequently associated with severe pathogenesis in the lower respiratory tract, such as bronchiolitis or pneumonia [62–68].

Animal-to-human (or vice versa) transmissions of SARS-CoV-2 can occur and can easily be considered secondary events [68]. Recent conclusions state that, so far, little is known concerning the role of pets and other animals (such as hamsters, minks, ferrets, lions, monkeys, tigers, and some others) in the transmission of COVID-19 [68, 69]. Thus, “as of now, there is no strong evidence for natural animal-to-human transmission or sustained animal-to-animal transmission of SARS-CoV-2” [68] and “animal-to-human transmission events of SARS-CoV-2 are considered rare” [69].

With the information available to date, there is no direct evidence to suggest which hypothesis is true: animal-to-human transmission of SARS-CoV-2 (recent or less recent, with or without the involvement of intermediate hosts); or an endemism of SARS-CoV-2, as proposed in this paper. The direct evidence only informs as to how viruses, including SARS-CoV-2, are related in a hierarchical classification.

On taxonomy, classification, naming, and binomial nomenclature of the coronaviruses

Binomial nomenclature is the designation of living organisms species' names in two Latin words: the first is the name of the genus, the second is the specific epithet (the name of the species). Ideally, the names of all virus species should be binomial.

However, despite intensive discussions [1–7], the names of virus species have consistently remained non-binomial, even within the taxonomic statements of ICTV or CSG and comprehensive virological reviews and theoretical studies [3, 55, 56, 70](Figure 1, Table S1), including those focused on the recognition of virus species [70].

Simultaneously, the current circumscriptions of the two largest genera of the *Coronaviridae* (*Alphacoronaviruses* and *Betacoronaviruses*) are very complicated. For example, current genus *Alphacoronavirus* circumscribes 12 subgenera, and current genus *Betacoronavirus* circumscribes five subgenera.

Widespread use of non-binomial names, as well as the acceptance of a complicated sub-generic structure of the current genera of coronaviruses, cause issues with the clear naming of viruses in the phylogenetic trees of *Coronaviridae*, as well as with the reading of these trees, especially by non-specialists.

Herein we resolved these issues by using the ICTV summaries [1–7] of the subgeneric names of *Coronaviridae* as the basic units of our notation, where possible, in the analyses and trees. However, we partly avoided the nomenclature methods traditionally used by virologists, instead demonstrating the possibilities of a simple and clear binomial nomenclature of the coronavirus family (Figures 2, S1–S4).

Accepting traditional monophyletic genera of the family at the rank of a tribe and, simultaneously, the current subgenera (that all are also monophyletic) at the generic rank is an easy, heuristic way to incorporate the Linnaean principles of the binomial nomenclature to the classification of the *Coronaviridae* literally "in one step" (Figure 2).

This solution also implies the future recognition of numerous undiscovered or unrecognized viral species, particularly in the currently monotypic subgenus *Sarbecovirus* [3]. For instance, the *intra*-species variation within the severe acute respiratory syndrome-related coronavirus (SARS-CoV), the current single species of subgenus *Sarbecovirus* (the same-name clade, subgenus *Betacoronavirus*) is actually comparable to *inter*-species variation within the other clades/subgenera of coronaviruses (e. g., within current subgenus/clade *Merbecovirus*) (Table S2).

As a point for discussion, based on these suggestions, it may be valid to treat the newly discovered "severe acute respiratory syndrome-related coronavirus 2" as a prospective new species of *Sarbecovirus*. In doing so, the coronavirus SARS-CoV-2 might be named a *Sarbecovirus* sp. (*Sarbecovirus* species), where "sp." or "species" is *any* available epithet (such as "*ambiguous*", "*vulgaris*" etc.). The trivial name "severe acute respiratory syndrome-related coronavirus 2" and correspondent abbreviation "SARS-CoV-2" could be listed in the description of *Sarbecovirus* sp. The same operations could be done with any other viruses of *Coronaviridae* (Figure 2).

Presently, numerous strains or variants of SARS-CoV-2 have been discovered [18, 72]. Because these strains all can be, in principle, established as forms or varieties of the same species (*Sarbecovirus* sp.), the special "dynamic nomenclature" for the SARS-CoV-2 strains [18] is unnecessary, even if it was practically useful during pandemic time.

Using such suggestions, the binary renaming of all of the species of the family, as well as other related viruses (either known or yet undiscovered), might not be a difficult task. In short, we believe that progress in virological taxonomy requires the complete rejection of the non-taxonomic category of "virus" in any future taxonomic studies. As in the well-developed traditional taxonomy of plants and animals, ranks such as "species", should be used instead.

Even from the conventional comparative standpoint, it is meaningless to lump two remarkably different viruses such as SARS and SARS-CoV-2 into the same species, the

"severe acute respiratory syndrome-related coronavirus" [70], in part due to the thousands of synapomorphic SNPs that clearly distinguish these two viruses from each other (Table S1). Such simple comparison may also be easily connected with the potential endemism of SARS-Cov-2 as well as with clear morphological differences between SARS-Cov and SARS-CoV-2 [71]. Thus, it seems to be more reasonable to establish the virus of severe acute respiratory syndrome-related coronavirus (SARS-CoV) under the binary name *Sarbecovirus* sp. 1 and the virus of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) under the binary name *Sarbecovirus* sp. X (e. g., *Sarbecovirus* sp. 6) (Figure 2, Table S1). This implies an increase of the rank of the current subgenus *Sarbecovirus* to the rank of genus and therefore the rank of current genus *Betacoronavirus* to the rank of tribe.

One may note that the viruses are the physical entities that therefore must be distinguished from species (concepts). In other words, a virus name such as "severe acute respiratory syndrome-related coronavirus 2" is not a synonym for a species name - but rather a name for a thing that belongs to an Aristotelian class. This suggestion is of course wrong because any trivial virus name (or its abbreviation, such as SARS-Cov-2) is still the name of the class that includes numerous individuals (virus particles).

The typological approach has not yet received proper implementation in virology, although some initial precedents are known [e.g., 73]. Additionally, for the purpose of discussion, we suggest consistent use of the Type Method within 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) be incorporated as a nomenclature type for any virus species. For example, the GenBank Accession number MN908947 could be treated as a nomenclature type of newly described SARS-CoV-2. Such a number implies the name/abbreviation of the biological isolate as well as other useful information.

If adopted, the higher nomenclature categories (tribes, genera, families, etc.) could be typified by the names of the species (genera, etc.), in exactly the same manner as plant or animal names. The nomenclature classification of plants and animals has been developed over hundreds of years, and as such is robust and well tested. Adopting the Linnaean binomial nomenclature for viruses will increase the universality of the system, and thereby lead to more consistent information content and information exchange across and within disciplines.

CONCLUSIONS

Methods of cladistics will be beneficial to virologists. Treating these methods within comparative and classification contexts, we have demonstrated that all four current genera of subfamily *Orthocoronavirinae* (family *Coronaviridae*), namely *Alphacoronavirus* (A), *Betacoronavirus* (B), *Deltacoronavirus* (Δ), and *Gammacoronavirus* (Γ), are monophyletic and representative of a *nature hierarchy* ((((A,B)Γ)Δ)) [74]. This hierarchy is, in principle, congruent to the general pattern of their hosts.

Using the same analytical approaches, we also confirmed that two subfamilies of *Coronaviridae*, namely monotypic subfamily *Letovirinae* and subfamily *Orthocoronavirinae*, are sisters, and as such, we would also confirm the accepting of *Letovirinae* as the previously proposed family *Abyssoviridae*.

All non-monotypic subgenera of all four genera of *Orthocoronavirinae* are monophyletic. Monotypic *Hibecovirus* is as a sister of {*Sarbecovirus*}. Bat coronaviruses RaTG13 or RmYN02, but not the pangolin coronavirus, are the closest relatives of SARS-CoV-2.

Accepting traditional genera of *Coronaviridae* at the rank of tribe and, simultaneously, the current subgenera of the same family at the generic rank, seems to be an easy and heuristic way to incorporate the Linnaean principles of the binnominal nomenclature to the classification of the family literally "in one step". For example, coronavirus SARS-CoV-2

might be named a *Sarbecovirus* species (*Sarbecovirus* sp.), where "species" (abbreviation "sp.") is any available epithet.

Any trivial virus name (or its abbreviation such as SARS-CoV-2) is still the name of the class that includes numerous individuals (virus particles). The non-taxonomic category "virus" must be eventually rejected in any future taxonomic studies in virology and the clear categories such as "species" must be used instead. This will aid future recognition of numerous virus species, particularly in monotypic [3] subgenus *Sarbecovirus*.

The Type Method must be widely used within the taxonomy of *Coronaviridae*, as well as any other virus family.

No member of any clade of *Coronaviridae* is an ancestor of any given virus species; the same is true regarding SARS-CoV-2. Humans are the only natural known host.

DECLARATIONS

Funding: This research did not receive any specific funding.

Conflict of interest: The authors declare that they have no conflicts of interests.

Availability of data and materials: Data are available upon request from the first author.

Code availability: No specific code has been developed during this study.

Authors contributions as per CRediT taxonomy: E.V.M.: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. M.L.T.: Investigation, Data Curation, Writing – original draft, Writing – review & editing, Supervision. N.E.M.: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Writing – review & editing, Visualization. L.S.: Investigation, Writing – review & editing. L.A.P.: Writing – review & editing. 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.

REFERENCES

1. Adams M.J., Lefkowitz E.J., King A.M.Q., Harrach B., Harrison R.L., Knowles N.J., Kropinski A.M., Krupovic M., Kuhn J.H., Mushegian A.R. et al. 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Arch. Virol.* 2017. V. 162. No. 5. P. 1441–1446. doi: 10.1007/s00705-016-3215-y
2. Walker P.J., Siddell S.G., Lefkowitz E.J., Mushegian A.R., Dempsey D.M., Dutilh B.E., Harrach B., Harrison R.L., Hendrickson R.C., Junglen S. et al. Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. *Arch. Virol.* 2019. V. 164. No. 9. P. 2417–2429. doi: 10.1007/s00705-019-04306-w
3. Gorbalenya A.E., Baker S.C., Baric R.S., de Groot R.J., Drosten C., Gulyaeva A.A., Haagmans B.L., Lauber C., Leontovich A.M., Neuman B.W. et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 2020. V. 5. No. 4. P. 536–544. doi: 10.1038/s41564-020-0695-z
4. Gorbalenya A.E., Krupovic M., Mushegian A., Kropinski A.M., Siddell S.G., Varsani A., Adams M.J., Davison A.J., Dutilh B.E., Harrach B. et al. The new scope of virus taxonomy, partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol.* 2020. V. 5. No. 6. P. 668–674. doi: 10.1038/s41564-020-0709-x
5. Siddell S.G., Walker P.J., Lefkowitz E.J., Mushegian A.R., Adams M.J., Dutilh B.E., Gorbalenya A.E., Harrach B., Harrison R.L., Junglen S. et al. Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of

Viruses (October 2018). *Arch. Virol.* 2019. V. 164. No. 3. P. 943–946. doi: 10.1007/s00705-018-04136-2

6. Siddell S.G., Walker P.J., Lefkowitz E.J., Mushegian A.R., Dutilh B.E., Harrach B., Harrison R.L., Junglen S., Knowles N.J., Kropinski A.M. et al. Binomial nomenclature for virus species, a consultation. *Arch. Virol.* 2020. V. 165. No. 2. P. 519–525. doi: 10.1007/s00705-019-04477-6
7. Siddell S.G., Walker P.J., Lefkowitz E.J., Mushegian A.R., Dutilh B.E., Harrach B., Harrison R.L., Junglen S., Knowles N.J., Kropinski A.M. et al. Correction to: Binomial nomenclature for virus species, a consultation. *Arch. Virol.* 2020. V. 165. No. 5. Article No. 1263. doi: 10.1007/s00705-020-04555-0
8. MacLachlan N.J., Dubovi E.J., Barthold S.W., Swayne D.F., Winton R.J. *Fenner's Veterinary Virology*. 5th edn. Amsterdam: Elsevier Inc, 2016.
9. Postler T., Clawson A.N., Amarasinghe G.K., Basler C.F., Bavari S., Benko M., Blasdell K.R., Briese T., Buchmeier M.J., Bukreyev A. et al. Possibility and challenges of conversion of current virus species names to Linnaean binomials. *Syst. Biol*. 2017. V. 66. No. 3. P. 463–473. doi: 10.1093/sysbio/syw096
10. Edgar R.C., Taylor J., Lin V., Altman T., Barbera P., Meleshko D., Lohr D., Novakovsky G., Buchfink B., Al-Shayeb B. et al. Petabase-scale sequence alignment catalyses viral discovery. *bioRxiv*. 2020. doi: 10.1101/2020.08.07.241729
11. Bukhari K., Mulley G., Gulyaeva A.A., Zhao L., Shu G., Jiang J., Neuman B.W. 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*. 2018. V. 524. P. 160–171. doi: 10.1016/j.virol.2018.08.010
12. Li T., Liu D., Yang Y., Guo J., Feng Y., Zhang X., Cheng S., Feng J. Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. *Sci. Rep*. 2020. V. 10. No. 1. Article No. 22366. doi: 10.1038/s41598-020-79484-8
13. Lu R., Zhao X., Li J., Niu P., Yang B., Wu H., Wang W., Song H., Huang B., Zhu N. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020. V. 395. No. 10224. P. 565–574. doi: 10.1016/S0140-6736(20)30251-8
14. Wu F., Zhao S., Yu B., Chen Y.-M., Wang W., Song Z.-G., Hu Y., Tao Z.-W., Tian J.-H., Pei Y.-Y. et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020. V. 579. No. 7798. P. 265–269. doi: 10.1038/s41586-020-2008-3
15. Kitching I.J., Forey P., Forey P.L., Humphries C., Williams D.M. *Cladistics, the Theory and Practice of Parsimony Analysis*. Oxford and New York: Oxford University Press, 1998. 228 p.
16. Williams D.M., Ebach M.C. *Foundations of Systematics and Biogeography*. New York: Springer, 2008. 310 p.
17. Williams D.M., Ebach M.C. *Cladistics, A guide to biological classification*. Cambridge: Cambridge University Press, 2020. 452 p.
18. Rambaut A., Holmes E.C., O'Toole Á., Hill V., McCrone J.T., Ruis C., du Plessis L., Pybus O.G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol*. 2020. V. 5. No. 11. P. 1403–1407. doi: 10.1038/s41564-020-0770-5
19. Farris J.S. The Logical Basis of Phylogenetic Analysis. In: *Advances in Cladistics*, *2*. Eds. Platnick N.I., Funk V. New York: Columbia University Press, 1983. P. 7–36.
20. Brower A.V.Z. Evolution is not a necessary assumption of cladistics. *Cladistics*. 2000. V. 16. No. 1. P. 143–154. doi: 10.1111/j.1096-0031.2000.tb00351.x
21. Mavrodiev E.V., Dell C., Schroder L. A laid-back trip through the Hennigian Forests. *PeerJ*. 2017. No. 5. Article No. e3578. doi: 10.7717/peerj.3578

22. Nelson G., Platnick N. *Systematics and Biogeography, Cladistics and Vicariance*. New York: Columbia University Press, 1981.
23. Nelson G., Platnick N. Three-taxon statements, a more precise use of parsimony? *Cladistics*. 1991. V. 7. No. 4. P. 351–366. doi: 10.1111/j.1096-0031.1991.tb00044.x
24. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol*. 1981. V. 17. No. 6. P. 368–376. doi: 10.1007/BF01734359
25. Felsenstein J. Inferring Phylogenies. Sunderland, MA: Oxford University Press; Sinauer Associates Inc., 2004.
26. Williams D.M., Ebach M.C., Wheeler Q.D. Beyond Belief: The Steady Resurrection of Phenetics. In: *Beyond Cladistics: The Branching of a Paradigm*. Eds. Williams D.M., Knapp S. Berkeley, CA: University of California Press, 2010. P. 169–196.
27. Tokarz R., Sameroff S., Hesse R.A., Hause B.M., Desai A., Jain K., Ian Lipkin W. Discovery of a novel nidovirus in cattle with respiratory disease. *J. Gen. Virol*. 2015. V. 96. No. 8. P. 2188–2193. doi: 10.1099/vir.0.000166
28. Wong A.C.P., Li X., Lau S.K.P., Woo P.C.Y. Global epidemiology of bat coronaviruses. *Viruses*. 2019. V. 11. No. 2. Article No. 174. doi: 10.3390/v11020174
29. Pipes L., Wang H., Huelsenbeck J.P., Nielsen R. Assessing uncertainty in the rooting of the SARS-CoV-2 phylogeny. *Mol. Biol. Evol.* 2021. V. 38. No. 4. P. 1537–1543. doi: 10.1093/molbev/msaa316
30. Zhou H., Chen X., Hu T., Li J., Song H., Liu Y., Wang P., Liu D., Yang J., Holmes E.C. 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. *Curr. Biol.* 2020. V. 30. No. 11. P. 2196–2203. doi: 10.1016/j.cub.2020.05.023
31. Zhou P., Yang X.-L., Wang X.-G., Hu B., Zhang L., Zhang W., Si H.-R., Zhu Y., Li B., Huang C.-L. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020. V. 579. No. 7798. P. 270–273. doi: 10.1038/s41586-020-2012-7
32. Liu P., Jiang J.-Z., Wan X.-F., Hua Y., Li L., Zhou J., Wang X., Hou F., Chen J., Zou J., Chen J. Are pangolins the intermediate host of the 2019 novel coronavirus, SARS-CoV-2? *PLoS Pathog*. 2020. V. 16. Article No. e1008421. doi: 10.1371/journal.ppat.1008421
33. Helmy Y.A., Fawzy M., Elaswad A., Sobieh A., Kenney S.P., Shehata A.A. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J. Clin. Med.* 2020. V. 9. No. 4. Article No. 1225. doi: 10.3390/jcm9041225
34. Draker R., Roper R.L., Petric M., Tellier R. The complete sequence of the bovine torovirus genome. *Virus Res*. 2006. V. 115. No. 1. P. 56–68. doi: 10.1016/j.virusres.2005.07.005
35. Katoh K., Misawa K., Kuma K., Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002. V. 30. No. 14. P. 3059–3066. doi: 10.1093/nar/gkf436
36. Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*. 2013. V. 30. No. 4. P. 772–780. doi: 10.1093/molbev/mst010
37. Miller M.A., Pfeiffer W., Schwartz T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In: *Gateway Computing Environments Workshop* (*GCE 2010*): Proceedings of a Meeting Held 14 November 2010, New Orleans, Louisiana, USA. San-Diego, CA: IEEE and Curran Associates Inc., 2010. doi: 10.1109/GCE.2010.5676129
38. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol*. 2000. V. 17. No. 4. P. 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

39. Gouy M., Guindon S., Gascuel O. SeaView version 4, a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 2010. V. 27. No. 2. P. 221–224. doi: 10.1093/molbev/msp259
40. Krishnamurthy S.R., Wang D. Origins and challenges of viral dark matter. *Virus Res*. 2017. No. 239. P. 136–142. doi: 10.1016/j.virusres.2017.02.002
41. Lapointe F.J., Cucumel G. The average consensus procedure, combination of weighted trees containing identical or overlapping sets of taxa. *Syst. Biol*. 1997. V. 46. No. 2. P. 306–312. doi: 10.1093/sysbio/46.2.306
42. Lapointe F.J., Levasseur C. Everything You Always Wanted to Know About the Average Consensus, and More. In: *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life*. Ed. Bininda-Emonds O.R.P. Dordrecht, Boston, London: Springer – Kluwer Academic Publisher, 2004. P. 87–105.
43. Maddison W.P., Maddison D.R. Mesquite, a Modular System for Evolutionary Analysis. Version 3.70. *Mesquite project*. 2021. URL: http://www.mesquiteproject.org (accessed 07.11.2022).
44. Swofford D.L. *PAUP*. Phylogenetic Analysis using Parsimony* and Other Methods*. Sunderland, MA: Sinauer Associates, 2002.
45. Rambaut A., Drummond A.J. *FigTree*. Version 1.4.4 [Interne] Institute of Evolutionary Biology, University of Edinburgh. 2018. URL: http://tree.bio.ed.ac.uk/software/figtree/ (accessed 07.11.2022).
46. Rineau V., Zaragüeta R., Bardin J. Information content of trees, three-taxon statements, inference rules and dependency. *Biol. J. Linn. Soc.* 2021. V. 133. No. 4. P. 1152–1170. doi: 10.1093/biolinnean/blab046
47. Mavrodiev E.V., Madorsky A. TAXODIUM Version 1.0: a simple way to generate uniform and fractionally weighted three-item matrices from various kinds of biological data. *PLoS ONE.* 2012. V. 7. No. 11. Article No. e48813. doi: 10.1371/journal.pone.0048813
48. Williams D.M., Siebert D.J. Characters Homology and, Three-Item Statement Analysis. In: *Homology and Systematics, Coding Characters for Phylogenetic Analysis*. Eds. Scotland R.W., Pennington R.T. New York: Chapman & Hall, London, 2000. P. 183–208.
49. Creevey C.J. *Clann: Investigating Phylogenetic Information Through Supertree Analyses*. Version 3.0. [Internet] The Lab. of James McInerney, Manchester, UK. 2004. URL: http://chriscreevey.github.io/clann/ (accessed 07.11.2022).
50. Creevey C.J., McInerney J.O. Trees From Trees: Construction of Phylogenetic Supertrees Using Clann. In: *Bioinformatics for DNA Sequence Analysis*. Ed. Posada D. New York: Springer Humana Press, 2009. P. 139–161.
51. Zhou X., Shen X.X., Hittinger C.T., Rokas A. Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Mol. Biol. Evol*. 2018. V. 35. No. 2. P. 486–503. doi: 10.1093/molbev/msx302
52. Trifinopoulos J., Nguyen L.T., von Haeseler A., Minh B.Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016. V. 44. No. W1. P. 232–235. doi: 10.1093/nar/gkw256
53. Anisimova M., Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* 2006. V. 55. No. 4. P. 539-552. doi: 10.1080/10635150600755453
54. Ferron F., Debat H.J., Shannon A., Decroly E., Canard B. A N7-guanine RNA cap methyltransferase signature-sequence as a genetic marker of large genome, non-mammalian Tobaniviridae. *NAR Genom. Bioinform.* 2020. V. 2. No. 1. Article No. lqz022. doi: 10.1093/nargab/lqz022
55. Sachs J.D., Karim S.S.A., Aknin L., Allen J., Brosbol K., Colombo F., Barron G.C., Espinosa M.F., Gaspar V., Gaviria A. et al. Commission on lessons for the future from

the COVID-19 pandemic. *The Lancet*. 2022. V. 400. No. 10359. P. 1224–1280. doi: 10.1016/S0140-6736(22)01585-9

56. Holmes E.C., Goldstein S.A., Rasmussen A.L., Robertson D.L., Crits-Christoph A., Wertheim J.O., Anthony S.J., Barclay W.S., Boni M.F., Doherty P.C. et al. The origins of SARS-CoV-2: A critical review. *Cell.* 2021. V. 184. No. 19. P. 4848–4856. doi: 10.1016/j.cell.2021.08.017
57. Garry R.F. SARS-CoV-2 furin cleavage site was not engineered. *PNAS*. 2022. V. 119. No. 40. Article No. e2211107119. doi: 10.1073/pnas.2211107119
58. Liu S.L., Saif L.J., Weiss S.R., Su L. No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. *Emerg. Microbes Infect*. 2020. V. 9. No. 1 P. 505–507. doi: 10.1080/22221751.2020.1733440
59. Domingo J.L. An updated review of the scientific literature on the origin of SARS-CoV-2. *Environ. Res*. 2022. V. 215. P. 114–131. doi: 10.1016/j.envres.2022.114131
60. Chaw S.M., Tai J.H., Chen S.L., Hsieh C.H., Chang S.Y., Yeh S.H., Yang W.S., Chen P.J., Wang H.Y. The origin and underlying driving forces of the SARS-CoV-2 outbreak. *J. Biomed. Sci.* 2020. V. 27. No. 1. Article No. 73. doi: 10.1186/s12929-020-00665-8
61. Conceicao C., Thakur N., Human S., Kelly J.T., Logan L., Bialy D., Bhat S., Stevenson-Leggett P., Zagrajek A.K., Hollinghurst P. et al. The SARS-CoV-2 spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol.* 2020. V. 18. No. 12. Article No. e3001016. doi: 10.1371/journal.pbio.3001016
62. Enjuanes L., Zuñiga S., Castaño-Rodriguez C., Gutierrez-Alvarez J., Canton J., Sola I. Molecular basis of coronavirus virulence and vaccine development. *Adv. Virus Res*. 2016. V. 96. No. 8. P. 245–286. doi: 10.1016/bs.aivir.2016.08.003
63. Fehr A.R., Perlman S. Coronaviruses: An Overview of Their Replication and Pathogenesis. In: *Coronaviruses. Methods in Molecular Biology,* vol. 1282. Eds. Maier H., Bickerton E., Britton P. New York: Humana Press, 2015. P. 1–26. doi: 10.1007/978-1-4939-2438-7_1
64. Fielding B.C. Editorial: Human coronavirus research: 20 years since the SARS-CoV outbreak. *Front. Microbiol*. 2022. V. 13. P. 1035267. doi: 10.3389/fmicb.2022.1035267
65. Liu D.X., Liang J.Q., Fung T.S. Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (*Coronaviridae*). *Encyclopedia of Virology*. 2021. V. 2. P. 428–440. doi: 10.1016/B978-0-12-809633-8.21501-X
66. van der Hoek L. Human coronaviruses, what do they cause? *Antivir. Ther*. 2007. V. 12. P. 651–658.
67. Woo P.C., Lau S.K., Chu C.M., Chan K.H., Tsoi H.W., Huang Y., Wong B.H., Poon R.W., Cai J.J., Luk W.K. et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol*. 2005. V. 79. No. 2. P. 884–895. doi: 10.1128/JVI.79.2.http:4-895.2005
68. Abdel-Moneim A.S., Abdelwhab E.M. Evidence for SARS-CoV-2 infection of animal hosts. *Pathogens*. 2020. V 9. No. 7. Article No. 529. doi: 10.3390/pathogens9070529
69. Murphy H., Ly H. What are the risk levels of humans contracting SARS-CoV-2 from pets and vice versa? *J. Med. Virol*. 2022. V. 94. No. 11. P. 5613–5614. doi: 10.1002/jmv.28035
70. Gorbalenya A.E., Siddell S.G. Recognizing species as a new focus of virus research. *PLoS Pathog*. 2021. V. 17. No.. 3. Article No. e1009318. doi: 10.1371/journal.ppat.1009318
71. Rossi G.A., Sacco O., Mancino E., Cristiani L., Midulla F. Differences and similarities between SARS-CoV and SARS-CoV-2: spike receptor-binding domain recognition and host cell infection with support of cellular serine proteases. *Infection*. 2020. V. 48. No. 5. P. 665–669. doi: 10.1007/s15010-020- 01486-5

72. Hadfield J., Megill C., Bell S.M., Huddleston J., Potter B., Callender C., Sagulenko P., Bedford T., Neher R.A. *Nextstrain*: real-time tracking of pathogen evolution. *Bioinformatics*. 2018. V. 34. No. 23. P. 4121–4123. doi: 10.1093/bioinformatics/bty407
73. Chaley M.B., Kutyrkin V.A. Coronavirus genus recognition based on prototype virus variants. *Math. Biol. Bioinf*. 2022. V. 17. No. 1. P. 10–27. doi: 10.17537/2022.17.10
74. Van Regenmortel M.H.V. Classes, taxa and categories in hierarchical virus classification: a review of current debates on definitions and names of virus species. *Bionomina*. 2016. V. 10. No. 1. P. 1–21. doi: 10.11646/bionomina.10.1.1

Received 03.08.2022.
Revised 06.11.2022.
Published 16.11.2022.