JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

Phylogenetic Tree Construction to Reveal the Detailed Evolution of SARS-CoV-2

Alaa Khudair Abbas Al-Khafaji, Bashar Talib Al-Nuaimi

Department of Computer Science, College of Science, University of Diyala, Iraq
scicompms2117@uodiyala.edu.iq

A B S T R A C T

The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), has spread worldwide. Therefore, this study aimed to create a phylogenetic tree construction of SARS-CoV-2 and other species of coronavirus by using the Maximum Likelihood Estimate approach after Perform Multiple Sequence Alignment. This study utilized 34 isolates of SARS-CoV-2 as well as another (Alpha & Beta) coronavirus species retrieved from the GenBank database (National Center for Biotechnology Information) for this investigation. The current findings showed that A highly accurate and trustworthy phylogenetic tree was obtained by using the Maximum Likelihood Estimate approach to show the evolutionary relationships linking between SARS-CoV-2s and other coronavirus species and determined the SARS-CoV-2 variations.

Keywords: SARS-CoV-2, phylogenetic tree construction, maximum-likelihood, phylogenetic inference**,** phylogenetic analysis, COVID-19.

INTRODUCTION

In December 2019, a novel coronavirus (severe acute respiratory syndrome coronavirus-2 [SARS-CoV-2]) was discovered in Wuhan/China, and was responsible for the Coronavirus Disease 2019 (COVID-2019) (WHO situation report May 30, 2020)[1]. SARS-CoV-2 caused a worldwide Pandemic and rapidly spread throughout the world[2]. Among the SARS-CoV-2 virus distinguishing characteristics is that it spreads very quickly (such as influenza virus and hepatitis C virus). The virus's evolutionary analysis revealed this. SARS-CoV-2 genetic variety was observed in a short period as a result of genetic variance in new viruses from their common ancestry. Even if SARS-CoV-2 are genetically similar to other coronavirus species, they demonstrate significant variances in terms of epidemiology, pathogenicity, and host spectrum[3]. We propose to investigate SARS-CoV-2 in Phylogenetic Tree Construction in order to estimate and visualize the evolutionary relationships between those genomes. In the field of epidemiology, viruses' evolutionary analysis is an invaluable resource.

In reality, SARS-CoV and MERS-CoV that caused epidemics in the previous are well-known to cause severe diseases to infect humans, besides SARS-CoV-2. SARS-CoV-2 is a different branch on the phylogenic tree from SARS-CoV and MERS-CoV, even though all three viruses were classified as Beta coronaviruses by complete genome analysis[2], [4]. Obviously from the above, understanding the evolutionary relationships among the SARS-CoV-2s and other coronaviruses species would be of great interest.

A phylogenetic tree is used to make estimates about how these species are related. These trees link the species together and show how they evolved [5]. Using the genes that these coronavirus species have in common; a well-supported phylogenetic tree has been constructed for SARS-CoV-2 [6].

For the phylogenetic analysis of SARS-CoV-2 in this study, the maximum likelihood approach was being used, with the aim of determining the origin and evolution of SARS-CoV-2. One of the most widely used methods of statistical estimation is that of maximum likelihood[7], [8]. The maximum likelihood approach is more difficult to apply and necessitates a deeper understanding of the evolutionary models upon which they are built. Because the maximum likelihood

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

method is more complex, they require a large number of computational steps, and the number of steps increases rapidly with the number of sequences, they are limited to a small number of sequences. They can be performed on a supercomputer to examine a large number of sequences simultaneously[9]. The current evolutionary methods need to be rethought in order to be more accurate reflections of reality[10].

This article thus aims to explore the possibility of finding relationships among the nucleotide sequences of SARS-CoV-2 using the genes shared by these species and wishes to analyse them in order to forecast the best trees illustrating the sequences' evolutionary relationships and to obtain a well-supported phylogenetic tree by using maximum likelihood method.

Material And Methods

Dataset construction

The complete genome of (34) was divided into (14) Alpha, and (20) Beta coronaviruses were Summoned by the NCBI (National Center for Biotechnology Information) database, which is available on the 5 of April 2022[11].

Table. 1. Information about some SARS-COV-2 and Coronavirus (Alpha and Beta) genomes

ACCESSION	DEFINITION	Genome size(bp)	Host	Country	Collection date	Lineage
NC_045512	SARS-CoV-2	29903	Homo sapiens	China	Dec-2019	Beta Sarbecovirus
OM039644	SARS-CoV-2	29821	Homo sapiens	USA: California	2021-12-18	Beta Sarbecovirus
KF514420	SARS coronavirus ExoN1	29687	lab_host="VeroE6 cells	USA: Nashville, TN	19-May-2009	Beta Sarbecovirus
MG772934	Bat SARS-like coronavirus	29732	Rhinolophus pusillus	China	Jul-2015	Beta Sarbecovirus
MT350598	Rousettus bat coronavirus GCCDC1	30162	Eonycteris spelaea	Singapore	Oct-2016	Beta Nobecovirus
MK967708	MERS-related coronavirus	30106	Camel	Egypt	2018	Beta Merbecovirus
MH940245	Human coronavirus HKU1	29811	Homo sapiens	Thailand	04-Jun-2017	Beta Embecovirus
MW386982	Porcine epidemic diarrhea virus	28052	Sus scrofa; breed: Gannan Black Small-eared Pig	China	Dec-2020	Alpha

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

MT438700	Human coronavirus 229E	27307	Homo sapiens	USA: Arkansas	27-Jan-2017	Alpha
MT444152	Feline coronavirus	28506	Cat	China	25-May-2019	Alpha

Construction of phylogenetic tree using complete genome sequences

We have followed a general flow chart consisting of several steps to construct a phylogenetic tree of Coronavirus and SARS-CoV-2 using maximum likelihood approaches as shown in Figure1.

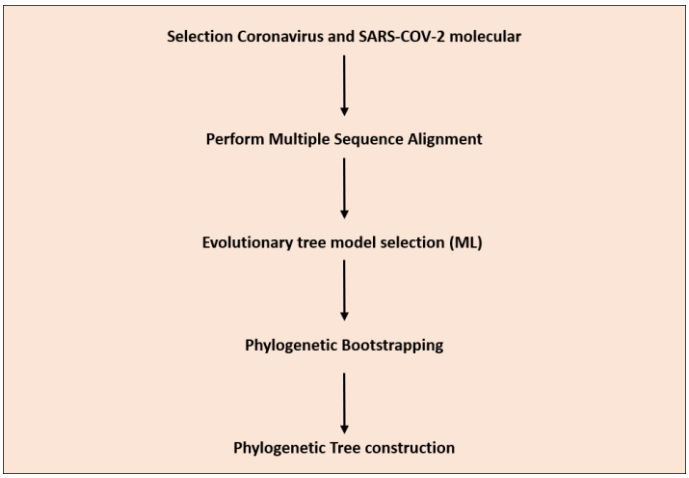


Figure.1 General steps of Coronavirus phylogenetic tree construction

Selection Coronavirus molecular

The first step is Coronavirus and SARS-CoV-2 molecular data Selection, It is possible to extract this from nucleotides or proteins sequence data. We used the nucleotide sequence complete genome while studying. We retrieved genome sequences from NCBI by accession number directly[12].

The SARS-CoV-2 is a virus that belongs to the coronavirus family (Coronaviridae). The Coronaviridae family has a large number of viral species, relatively. Letovirinae and Orthocoronavirinae are subfamilies of the coronavirus family. SARS-CoV-2 belongs to the subfamily of orthocoronaviruses. Orthocoronaviruses are further classified into four genera: Alpha, Beta, Gamma, and Delta coronaviruses .Beta coronaviruses are classified into five subgenus: Embecovirus (lineage A), Sarbecovirus (lineage B), Merbecovirus (lineage C), and Nobecovirus (lineage D)[13], [14]. According to this categorization, SARS-CoV and SARS-CoV-2 (one of them is NC_045512 Reference Sequence the first SARS-CoV-2 isolate sequencing discovered in Wuhan/China and was dated 2019-12-20 based on the appearance of the sickness in the patient [15]) are classified in Sarbecovirus. Despite the fact that SARS-CoV-2 is an (RNA) virus, the nucleotide sequences that have been deposited are in DNA form. Table1 contains information on these genomes. Those genomes of SARS-CoV and coronaviruses species that resembled the SARS-CoV-2 genomes were chosen. The sequences' integrity was verified. And databases were created by identifying genomic sequences with the country, host, and date of collection, as well as other information.

Perform Multiple Sequence Alignment

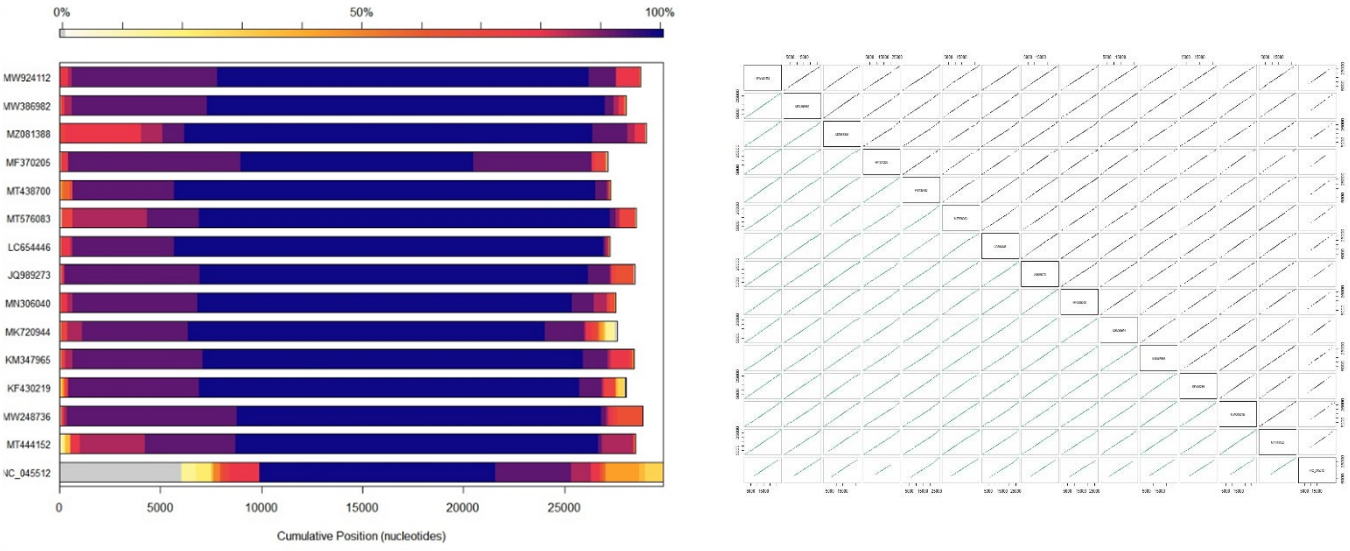
After the selection Coronavirus data stage, the sequences retrieved must be aligned [16], [17]. Determining a common beginning point for these circular genomes is the initial stage in the alignment process[10]. Sequence alignment is the process of rearranging DNA, RNA, or protein sequences in order to identify areas of similarity that may be used to align the complete sequence. An evolutionary relationship may be determined by the similarity between the sequences [18].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

Multiple Sequence Alignment (MSA) aligns three or more protein or nucleotide sequences and computes a consensus sequence. MSA is simple when all the sequence is similar, but challenging if they are not. If so, a large number of gaps will be required[19]. Multiple alignments are an important stage because they demonstrate positional symmetry in sequences evolution. Only a succeeded and accurate sequence alignment leads to a tree that is genealogically interrelated[20]. Furthermore, It may assist in the detection of mutations or recombination events in pairs of closely related genomes[10].

Synteny will be used to compare genomes and will do pairwise analysis on each set of genomes supplied, by the FindSynteny() function [21]. In general, syntenic matching occurs when two genomes share the same regions. We thus obtained the first representation of synteny of the (14) Alpha coronavirus with SARS-CoV-2 Wuhan-Hu-1 (NC_045512.2) outgroup shown in Figure.2 and (20) Beta coronavirus genomes shown in Figure.3, we find syntenic blocks between sequences [21]. As can be seen, these (34) genomes exhibit a mixture of sequence similarity and divergence with relatively few recombination events. A lack of syntenic blocks can be seen in a grey region of the subject's genome, while dark blue indicates complete similarity. Rearrangement possibilities are indicated in certain locations by the ramp's inversion, while in other areas, the blocks are reordered from the default colour bar.

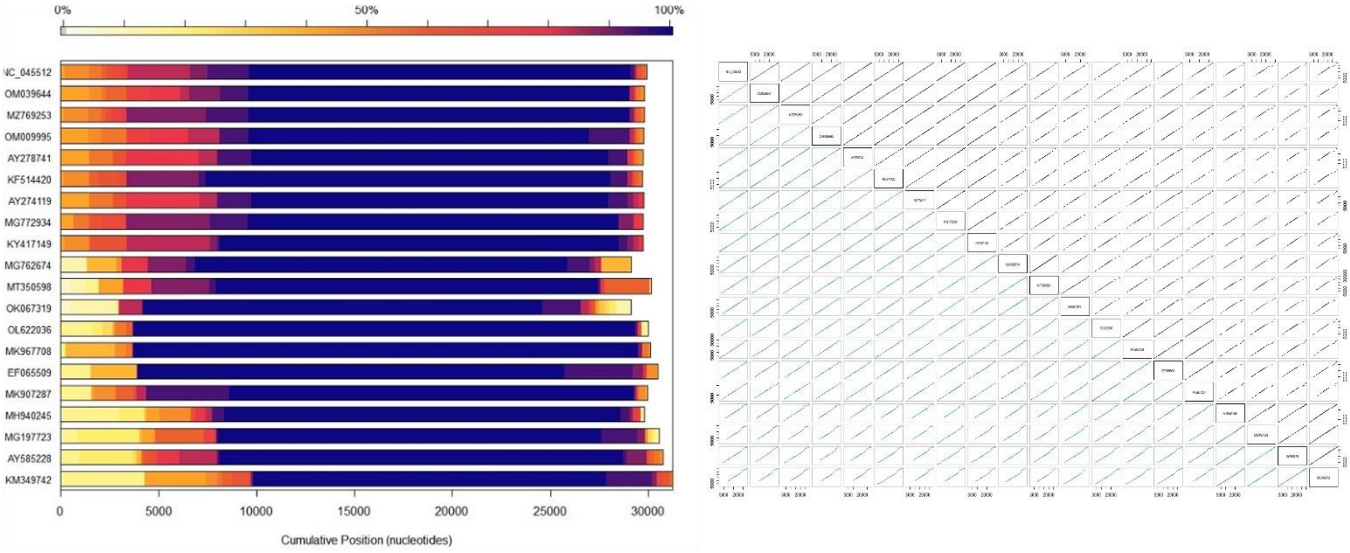
We perform multiple sequence alignment for the SARS-CoV-2 and other coronavirus' complete genome aligned by the *MSA* function in the msa R package using the ClustalW algorithms, which achieved a good accuracy[22]. Figure.4 shows a part of coronaviruses and SARS-CoV-2 multiple sequence alignment.



(a) **(b)**

Figure.2. (a) Adjacent pairs synteny blocks representation of Alpha coronavirus sequences. When the genome shares the same region with the first genome, their coloration is the same hue, or grey if they don't. **(b)** Dot plot supplies an alternate synteny map visualizing of Alpha coronavirus. Black color diagonal stripes represent syntenic areas having identical directions. All of these species have significant sequence similarities and minimal recombination rate.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452



(a)

(b)

Figure.3. (a) Adjacent pairs synteny blocks representation of Beta and SARS-CoV-2 coronavirus sequences. When the genome shares the same region with the first genome, their coloration is the same hue, or grey if they don't. (b) Dot plot supplies an alternate synteny map visualizing Beta and SARS-CoV-2 coronavirus. Black color diagonal stripes represent syntenic areas having identical directions. All of these species have significant sequence similarities and minimal recombination rate.

Evolutionary tree model selection

To construct phylogenetic trees, statistical approaches are utilized to find the best explain the evolutionary relationships of the aligned sequences in a data set[23]. The most often used computational approaches are distance-based approaches and character-based approaches such as maximum-parsimony and maximum-likelihood[23].

The Maximum Likelihood Estimate (MLE) is utilized to find the topologies and branches length that is the highest likelihood of producing the aligned data, hence giving the substitution models and tree. After the alignment stage, the likelihood value is calculated by evaluating several nucleotide substitutions models. [24], [25]. A quartet programme is used to fulfill the searching area. This latest procedure determines all potential sequence combinations for the purpose of tree reconstruction [25]. MLE are reported to be best under all circumstances [7].

MLE is the most often used approach to statistical inference because of its high consistency and asymptotical normal distribution[26]. The fundamental problem of these approaches is that they are computationally costly. However, with modern computers, this is less of a major issue.[25].

We used MLE approaches to construct a phylogenetic tree, which can incorporate all multiple sequence alignment data into a statistical framework for estimating model parameters. We compute the likelihood of a phylogenetic tree first. Then, we optimize the different model parameters and branch lengths for the selected model of nucleotide evolution which used GTR Gamma mode substitution models as recommended by JModelTest 2.1.10 to get optimized. The output tree was visualized utilizing the Interactive Tree of Life (iTOL) online tool[27].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

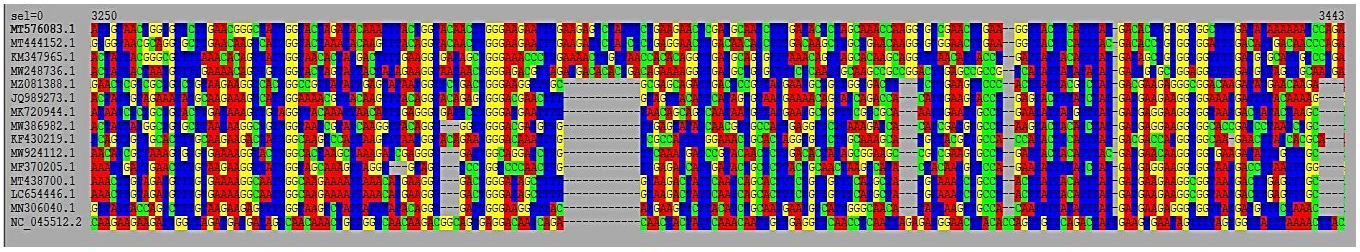
Phylogenetic Tree construction

Almost every statistical estimate may be evaluated for accuracy using the bootstrapping method, which is computer-based. Nonparametric estimate issues where analytic approaches are impracticable and commonly utilized can benefit from this technique [28]. Bootstrap values may also be affected artifactually by the total number of characters in the dataset [29].

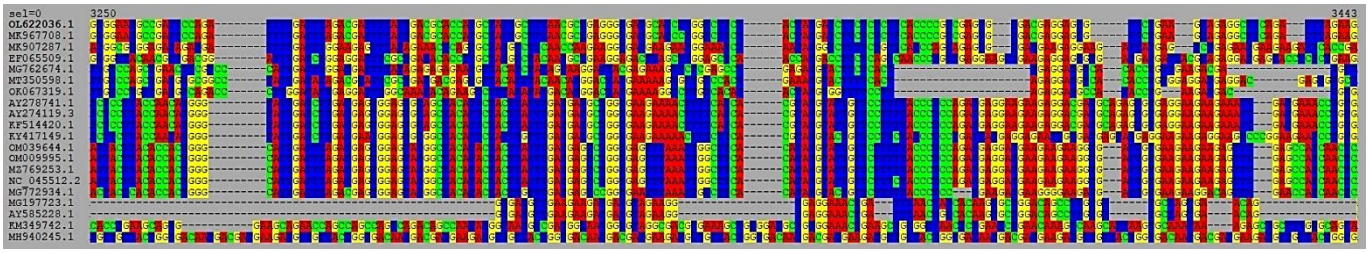
A phylogenetic tree representation contains a great deal of information, both explicit and implicit. Because some of the implicit information may be deceptive, it is the investigator's responsibility to guarantee that all information presented, explicit and implicit, is accurate[30]. It determines the evolutionary relationships among tips and nods.

Each phylogenetic tree node indicates the final common ancestor of the two lineages descending from the node's parent [31]. Inner branches join two nodes, while exterior branches join a tip and a node as seen in the illustration[31]. Using the branch lengths, we may estimate how much change has happened between any two nodes. A tree represent expresses this information well[30].

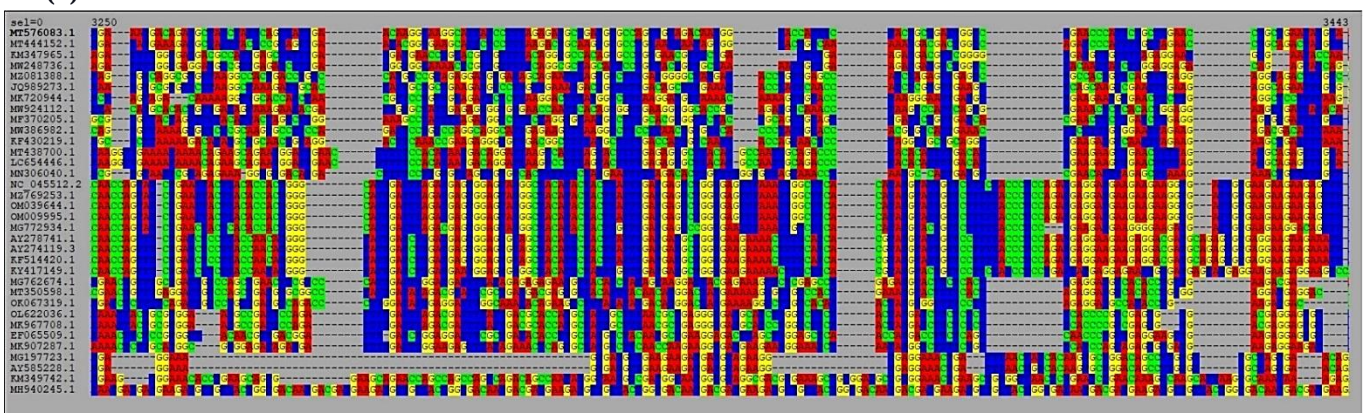
We used the plotBS() function which is a function in the phangorn R package that plots a phylogenetic tree with the bootstrap values assigned to them (internal) edges[32]. It can also use to assign bootstrap values to a phylogenetic tree. Also used the “phylogram” type in which inner nodes are depicted by vertical lines to construct the tree. Figure 5,6,7 shows a representation of the phylogenetic trees with bootstraps.



(a)



(b)



(c)

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

Figure.4. (a) MSA of Alpha coronavirus sequences with outgroup NC_045512.2 SARS-CoV-2 reference sequence. **(b)** MSA of Beta coronavirus sequences. **(c)** MSA of Alpha and Beta coronavirus sequences together.

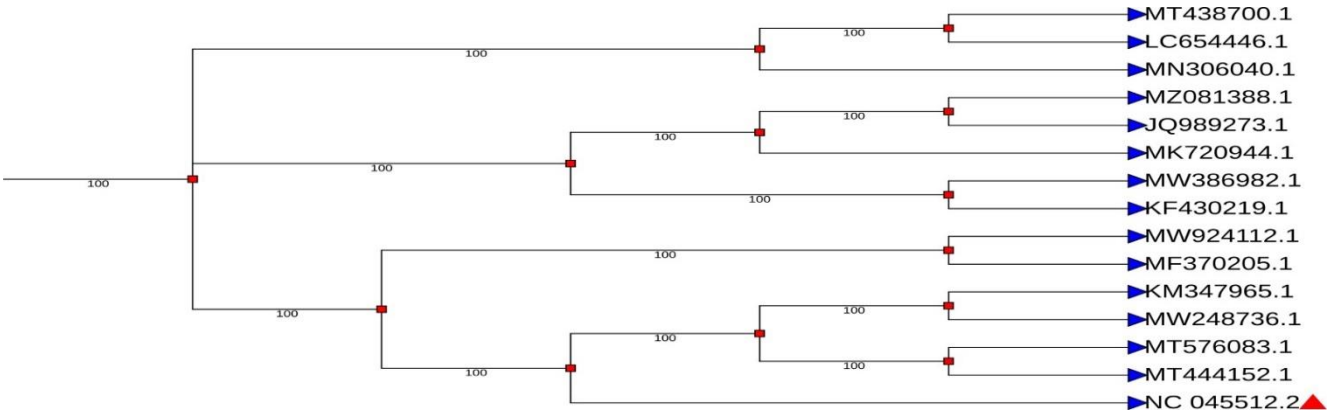


Figure.5 The phylogenetic tree of (14) Alpha coronaviruses using the Maximum Likelihood approach with 1000 bootstraps, the red triangle indicating SARS-CoV-2 RefSeq.

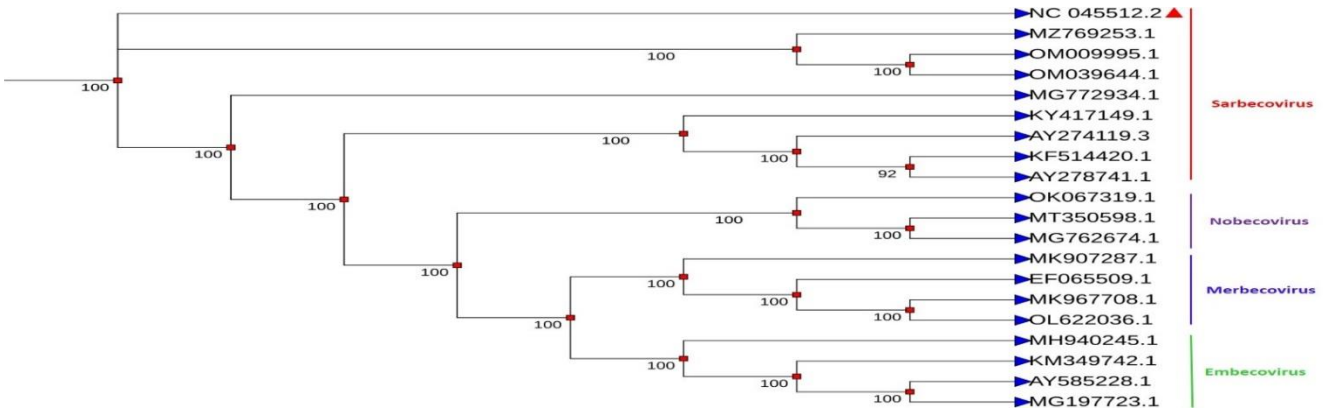


Figure.6 The phylogenetic tree of (20) Beta coronaviruses by Maximum Likelihood approach with 1000 bootstraps, the red triangle indicating SARS-CoV-2 RefSeq.

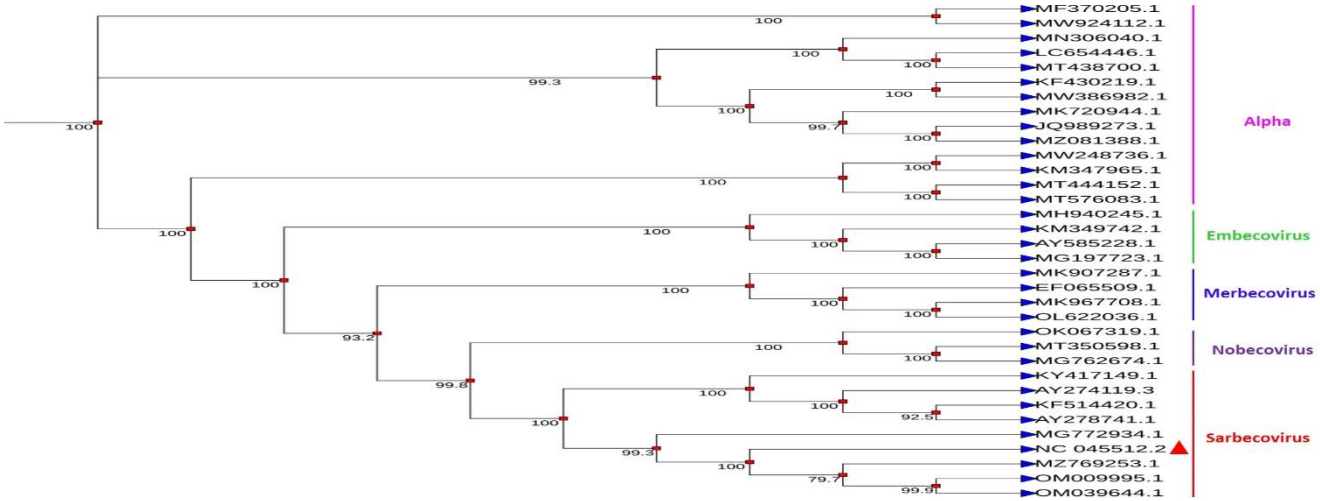


Figure.7 The phylogenetic tree of (14) Alpha coronaviruses and (20) Beta coronaviruses by ML method with 1000 bootstraps, the red triangle indicating SARS-CoV-2 RefSeq.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

RESULTS AND DISCUSSION

We have obtained three phylogenetic trees based on complete genome sequences using the maximum likelihood estimate (MLE) approach, the first is for Alpha coronavirus species only shown in (Figure.5), the second is for Beta coronavirus species, including SARS-COV-2 Species only as shown in (Figure.6), and the third tree is constructed from merging Alpha and Beta phylogenetic tree together and as shown in (Figure.7). we removed samples that are too divergent from others. Despite the close relationship among these organisms, the genome length and content are very different, As shown in (Table 1).

The resulting phylogenetic tree topologies vary dramatically in structure, owing to differences stemming from genetic diversity and frequent mutations in the expanding Alpha, beta coronavirus, and SARS-COV-2 Species genome.

Alpha phylogenetic tree

We choose our Alpha coronavirus sequences from this controversial dataset using the complete genome from NCBI encompassing: Alphacoronavirus HCQD-2020 (MW924112), Porcine epidemic diarrhea virus (MW386982), Alphacoronavirus sp (MZ081388), Rhinolophus bat coronavirus HKU2 (MF370205), Human coronavirus 229E (MT438700), Transmissible gastroenteritis virus (MT576083), Human coronavirus 229E Fukushima_H832_2020 (LC654446), Bat-coronavirus (KF430219), Hipposideros bat coronavirus HKU10 (JQ989273),Human coronavirus NL63 (MN306040), Tylonycteris bat coronavirus HKU33 (MK720944), Ferret coronavirus (KM347965), Lucheng-Rn-rat coronavirus (NC_032730), MAG: Minacovirus (MW248736) and Feline coronavirus (MT444152).

For all the 14 Alpha coronavirus genomes with SARS-CoV-2 Wuhan-Hu-1 (NC_045512) outgroup shown in Figure.2, the *MSA* function has been used to align them as previously stated. Figure2 shows the initial representation of synteny for all of them. Alpha coronavirus genomes are remarkably similar and match each other perfectly, as can be observed, Furthermore, just a few recombination events have taken place inside these genomes. We also see that some regions in the outgroup genome appeared in grey.

The obtained tree is well-supported (cf. Figure.5). And the obtained tree is trustworthy. We may deduce these genomes are extremely preserved since they contain regions that are almost similar to one another and have undergone no rearrangement.

Beta phylogenetic tree

We retrieved complete genomes of the (20) Beta coronavirus sequence available in NCBI directly, namely by lineage: Sarbecovirus(9 genomes), Embecovirus(4 genomes), Merbecovirus(4 genomes), Nobecovirus(3 genomes).

We choose our Beta coronavirus sequences from this controversial dataset from NCBI encompassing: Embecovirus lineage including Human coronavirus HKU1 isolate SI17244 (MH940245), Human coronavirus OC43 strain HZ-459 (MG197723), Human coronavirus OC43 strain ATCC-VR-759 (AY585228), Betacoronavirus HKU24 (KM349742), Merbecovirus lineage include MERS related coronavirus isolate MERS-CoV_Dammam_2019 (OL622036), MERS-related coronavirus isolate Merscov/Egypt/Camel/AHRI-FAO-1/2018 (MK967708), Pipistrellus bat coronavirus HKU5 (EF065509), Erinaceus hedgehog coronavirus HKU31 (MK907287), Nobecovirus lineage include Rousettus bat coronavirus HKU9 (MG762674), Rousettus bat coronavirus GCCDC1 (MT350598) and Pteropus Rufus nobecovirus (OK067319).

And the complete genome of the Sarbecovirus lineage in this dataset, Which we have chosen encompasses: SARS-COV-2 isolate Wuhan-Hu-1 Reference Sequence (NC_045512.2), SARS-COV-2 isolate SARS-CoV-2/human/USA/CA-CDC-FG-202336/2021 (OM039644), SARS-COV-2 isolate SARS-CoV-2/human/USA/CO-CDPHE-2102267526/2021 (OM009995), SARS-COV-2 isolate SARS-CoV-2/human/BHR/342425280/2021 (MZ769253), SARS coronavirus Urbani

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 538 - 549
https://publishoa.com
ISSN: 1309-3452

(AY278741), SARS coronavirus ExoN1 (KF514420), SARS coronavirus Tor2 (AY274119), Bat SARS-like coronavirus isolate bat-SL-CoVZXC21 (MG772934) and Bat SARS-like coronavirus isolate Rs4255 (KY417149).

Following these operations, it is possible to obtain a synteny representation of Beta coronavirus sequences, as seen in Figure3. There are a few variances almost certainly reflect divergent evolutionary origins. and are very similar, showing homologous regions among many Beta genomes.

The maximum likelihood tree was obtained by using Beta coronavirus's complete genome in conjunction with the GTR Gamma substitution model. Figure6 illustrates this, all branches show a well-supported, Except for what is exhibited between SARS coronavirus Urbani (AY278741) and SARS coronavirus ExoN1 (KF514420) 92% bootstrap support value.

The phylogenic tree that was constructed showed that the SARS-COV-2 locate in the Sarbecovirus lineage, Beta coronavirus, and is associated with the SARS-COV as well as Bat-SARS-like coronavirus. Three main branches appeared in the phylogenetic tree, it appears in the first branch SARS-COV-2 isolates Wuhan- Hu-1 (NC_045512.2), which is a reference sequence.

As for the rest of the sequences in the third branch, all the lineage sequences were grouped into a group and a very well-supported phylogenetic tree has been obtained.

Bootstrapping value between major phylogenetic tree branches indicates that there is a great match in the composition and content of their genomes in general. The group of the second major branch indicates the great similarity between the SARS-COV-2 sequences, so they are collected in one branch.

The proximity of the Bat SARS-like coronavirus and SARS coronavirus sequences to the SARS-COV-2s sequences in the phylogenetic tree was due to their genetic affinity and similarity and that they are descended from the same ancestor.

Alpha and Beta phylogenetic tree

Phylogenetic trees are being conducted extensively in order to better understand the genomic properties of SARS-CoV-2 [33]. The Construction Phylogenetic Tree for Alpha and Beta Coronaviruses showed the evolutionary relationships among SARS-COV-2s and alpha Coronaviruses, Beta Coronaviruses sequences. The SARS-CoV-2 sequences showed no high differences between them and other sequences of a phylogenetic tree by the Maximum Likelihood method (Figure.7). SARS-CoV-2 is clustered with Beta coronavirus genera, as the evolutionary tree demonstrates. Furthermore, SARS-CoV-2 belongs to a distinct clade than other alpha and beta Coronavirus sequences. With a high bootstrap value, convergent evolution appears to be a possibility (Figure7). High bootstrap values between 79.7 and 100 ensure the trustworthiness of the phylogenetic tree of Alpha and Beta coronavirus in the maximum likelihood estimation.

The SARS-CoV-2 sequences group has been shown to be 99.9% identical among SARS-COV-2 (OM039644) and (OM009995) and found to be 79.7% identical between them and SARS-COV-2 (MZ769253). Also, we found it to be 100% identical between the Wuhan-Hu-1 RefSeq (NC_045512) and other sequences of SARS-CoV-2. The SARS-CoV-2 sequences group and Bat SARS-like coronavirus were 99.3% similar (MG772934), Those who are descended from one node (ancestor).

We found 99.8% identical between Sarbecovirus lineage sequences which include SARS-CoV-2 sequences and Nobecovirus lineage sequences), Those who are descended from one node (ancestor).

We found 93.2% identical between the group of Sarbecovirus lineage sequences (which include SARS-CoV-2 sequences), Nobecovirus lineage sequences, and Merbecovirus lineage sequences which include (MERS related coronavirus and Pipistrellus bat coronavirus HKU5), Those who are descended from one node (ancestor). And all of them

which are identical to Embecovirus lineage sequences including (Human coronavirus HKU1 and Human coronavirus OC43) are well-supported.

We also found four sequences of the alpha coronavirus tree which included MAG: Minacovirus (MW248736), Transmissible gastroenteritis virus (MT576083), Ferret coronavirus (KM347965), and Feline coronavirus (MT444152) share a single ancestor (node) with the beta coronavirus sequences and the bootstrap value was well-supported. This was due to the great convergence between this group sequence of the alpha coronavirus tree and the sequences of the beta coronavirus.

The difference in bootstrapping value between the tree in (Figure.6) and the tree in (Figure 7) for the same sequences as it appeared between the sequences (OM039644) and (OM009995) which were in (Figure.6) was 100%, but in (Figure.7) it appeared 79.7% was due to the gaps through a multiple sequence alignment process. This confirms that there are mutations that have occurred in SARS-CoV-2 sequences and the Omicron virus was the result of these changes. Bootstrap values may also be affected artifactually by the total number of characters in the data set.

We have shown from the results that the maximum likelihood approach is efficient and accurate in constructing a phylogenetic tree for coronavirus species, including SARS-CoV-2 to show the evolutionary relationships.

Conclusion

In this article, we constructed a phylogenetic tree of Alpha and Beta coronavirus to show the evolutionary relationships between SARS-CoV-2 and other species of coronavirus by using the MLE approach. A highly accurate and trustworthy phylogenetic tree was obtained.

The development of variants in the COVID-19 pandemic is probable and may occur quickly. As a result, strains of SARS-CoV-2 should be continually monitored. For monitoring the evolution of SARS-CoV-2s lineage, the phylogenetic tree and the MLE approach appear to be effective and beneficial due the obtained tree is trustworthy.

In future work, we intend to a construct phylogenetic tree of SARS-CoV-2s to show the evolutionary relationships with Alpha, Beta, Gumma, and Delta coronavirus, revealing a variety of SARS-CoV-2s mutations and single nucleotide polymorphisms (SNPs). In order to control and treat COVID-19 and avoid another epidemic, it is critical that the variety and development of SARS-CoV-2 be closely monitored at all times.

REFERENCES

[1] B. Morel *et al.*, “Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult,” *Molecular Biology and Evolution*, vol. 38, no. 5, pp. 1777–1791, May 2021, doi: 10.1093/molbev/msaa314.

[2] T. Li *et al.*, “Phylogenetic supertree reveal detailed evolution of SARS-CoV-2,” *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-79484-8.

[3] M. Sallam and A. Mahafzah, “Molecular analysis of sars-cov-2 genetic lineages in Jordan: Tracking the introduction and spread of covid-19 UK variant of concern at a country level,” *Pathogens*, vol. 10, no. 3, pp. 1–12, Mar. 2021, doi: 10.3390/pathogens10030302.

[4] N. Zhu *et al.*, “A Novel Coronavirus from Patients with Pneumonia in China, 2019,” *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/nejmoa2001017.

[5] J. Rizzo and E. C. Rouchka, “Review of Phylogenetic Tree Construction,” 2007. [Online]. Available: www.docsity.com

[6] B. Al-Nuaimi, B. Alkindy, J.-F. Couchot, M. Salomon, and C. Guyeux, "Ancestral Reconstruction and Investigations of Genomic Recombination on Campanulides Chloroplasts.", Journal of Integrative Bioinformatics, 2019.

[7] L. le Cam, "Maximum Likelihood: An Introduction," International Statistical Review / Revue Internationale de Statistique 1990.

[8] I. J. Myung, "Tutorial on maximum likelihood estimation," *Journal of Mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003, doi: 10.1016/S0022-2496(02)00028-7.

[9] David W."Bioinformatics sequence and genome analysis". http://www-nbrf.georgetown.edu/pir,2001.

[10] C. Guyeux, B. Al-Nuaimi, B. AlKindy, J.-F. Couchot, and M. Salomon, "On the ability to reconstruct ancestral genomes from Mycobacterium genus," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1705.00276

[11] "NCBI Virus." https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/ (accessed Apr. 05, 2022).

[12] "Package 'traits' Title Species Trait Data from Around the Web," 2020. [Online]. Available: https://docs.ropensci.org/traits/,

[13] S. Nakagawa and T. Miyazawa, "Genome evolution of SARS-CoV-2 and its virological characteristics," *Inflammation and Regeneration*, vol. 40, no. 1. BioMed Central Ltd, Aug. 10, 2020. doi: 10.1186/s41232-020-00126-7.

[14] A. Tabibzadeh *et al.*, "Evolutionary study of COVID-19, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as an emerging coronavirus: Phylogenetic analysis and literature review," *Veterinary Medicine and Science*, vol. 7, no. 2, pp. 559–571, Mar. 2021, doi: 10.1002/vms3.394.

[15] N. M. H. Phan, H. Faddy, R. Flower, K. Spann, and E. Roulis, "Ancestral Area Reconstruction of SARS-CoV-2 Indicates Multiple Sources of Entry into Australia," *The Open Bioinformatics Journal*, vol. 14, no. 1, pp. 13–20, Nov. 2021, doi: 10.2174/1875036202114010013.

[16] C. Kemena and C. Notredame, "Upcoming challenges for multiple sequence alignment methods in the high-throughput era," *Bioinformatics*, vol. 25, no. 19. pp. 2455–2465, 2009. doi: 10.1093/bioinformatics/btp452.

[17] C. Guyeux, B. Al-Nuaimi, B. AlKindy, J. F. Couchot, and M. Salomon, "On the reconstruction of the ancestral bacterial genomes in genus Mycobacterium and Brucella," *BMC Systems Biology*, vol. 12, Nov. 2018, doi: 10.1186/s12918-018-0618-2.

[18] S. Vijayakumar, A. Bhargavi, U. Praseeda, and S. A. Ahamed, "Optimizing sequence alignment in cloud using hadoop and MPP database," in *Proceedings - 2012 IEEE 5th International Conference on Cloud Computing, CLOUD 2012*, 2012, pp. 819–827. doi: 10.1109/CLOUD.2012.34.

[19] M. Dipl, I. C. Horejš, H. Horejš-Kainrath, U. Bodenhofer, and J. Kepler, "Multiple Sequence Alignment with R," 2016. [Online]. Available: www.jku.at

[20] S. R. Amit Roy, "Molecular Markers in Phylogenetic Studies-A Review," *Journal of Phylogenetics & Evolutionary Biology*, vol. 02, no. 02, 2014, doi: 10.4172/2329-9002.1000131.

[21] Wright, E. "DECIPHER". Available from: https://git.bioconductor.org/packages/DECIPHER (Accessed on 03.05.2022),2021.

[22] E. Bonatesta, C. Kainrath, and U. Bodenhofer, "msa An R Package for Multiple Sequence Alignment Software Manual," 2021.

[23] S. Aryal, "How to construct a Phylogenetic tree ?," Available from: https://thebiologynotes.com/how-to-construct-a-phylogenetic-tree/, (Accessed on 05.04.2022),2019.

[24] L. Addario-Berry, B. Chor, M. Hallett, J. Lagergren, A. Panconesi, and T. Wareham, "ANCESTRAL MAXIMUM LIKELIHOOD OF EVOLUTIONARY TREES IS HARD,", Journal of Bioinformatics and Computational Biology, 2004. [Online]. Available: www.worldscientific.com

[25] B. Talib, H. Al-Nuaimi, and P. C. Guyeux, "Ancestral Reconstruction and Investigations of Genomics Recombination on Chloroplasts Genomes," Journal of integrative bioinformatics,2017.

[26] Pan, Fang. "Growth Curve Models and Statistical Diagnostics, " Springer Series in Statistics,2002.

[27] I. Letunic and P. Bork, "Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation," *Nucleic Acids Research*, vol. 49, no. W1, pp. W293–W296, Jul. 2021, doi: 10.1093/NAR/GKAB301.

[28] B. Efron, E. Halloran, and S. Holmes, "Bootstrap confidence levels for phylogenetic trees," Proc. Natl.Acad. Sci. USA 1996.

[29] P. S. Soltis and D. E. Soltis, "Applying the Bootstrap in Phylogeny Reconstruction," Statistical Science, 2003.

[30] B. G. Hall, "Building phylogenetic trees from molecular data with MEGA," *Molecular Biology and Evolution*, vol. 30, no. 5, pp. 1229–1235, May 2013, doi: 10.1093/molbev/mst012.

[31] D. Baum, "Reading a phylogenetic tree: The meaning of monophyletic groups," Nature Education, 2008. [Online]. Available: http://www.nature.com/scitable/topicpage/Reading-a-Phylogenetic-Tre...

[32] K. Schliep, A. J. Potts, D. A. Morrison, and G. W. Grimm, "Intertwining phylogenetic trees and networks," *Methods in Ecology and Evolution*, vol. 8, no. 10, pp. 1212–1220, Oct. 2017, doi: 10.1111/2041-210X.12760.

[33] S. B. Kadam, G. S. Sukhramani, P. Bishnoi, A. A. Pable, and V. T. Barvkar, "SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights," *Journal of Basic Microbiology*, vol. 61, no. 3. Wiley-VCH Verlag, pp. 180–202, Mar. 01, 2021. doi: 10.1002/jobm.202000537.