OPEN ACCESS Freely available online

Biological Systems: Open Access

Research Article

Phylogenetic Grouping and Levels of Genetic Structuration in SARS-Cov-2, Using FIGTREE V1.4.4 Software

Eduarda Doralice Alves Braz Da Silva[1]; Dallynne Bárbara Ramos Venancio[1]; Pierre Teodosio Felix[*1]; Robson da Silva Ramos[1]

[1]*Laboratory of Population Genetics and Computational Evolutionary Biology - LaBECom, UNIVISA, Vitoria de Santo Antao, Pernambuco, Brazil*

ABSTRACT

In this work, 38 complete genomes of SARS-CoV-2 were used, from five Central American countries (Belize, Guatemala, Cuba, Jamaica and Puerto Rico) with 04, 10, 2, 8 and 14 haplotypes, respectively, with an extension of up to 29,885 bp. All sequences were publicly available on the National Biotechnology Information Centre (NCBI) platform and were previously aligned with the MEGA X software, where all gaps and ambiguous sites were extracted for the construction of the phylogenetic tree. Only 79 sites remained preserved and the high number of polymorphisms helped to establish a clear pattern of non-genetic structuring, based on the time of divergence between the groups, visualized with the software FIGTREE V. 1.4.4.

Keywords: Bioinformatics, FIGTREE, Phylogeny, SARS-CoV-2, Coronavirus, Bio mathematical.

INTRODUCTION

With rapid person-person dissemination and vertiginous expansion in regions marked by insufficient sanitation and hygiene measures, SARS-CoV-2 has been establishing itself as a pandemic whose efforts in the use of protocols are still not efficient and the lack of medicines and vaccines, still in the testing phase worldwide, only increase the problem. (FELIX et al, 2020). Trying to understand evolutionary aspects of the virus, the team of the Laboratory of Population Genetics and Computational Evolutionary Biology (LaBECom-UNIVISA) performed a phylogenetic analysis with a structural bias in a PopSet of complete genomes of the SARS-CoV-2 virus from five Central American countries, available at the National Biotechnology Information Center (NCBI).

METHODS

Database: The 38 complete SARS-CoV-2 genomes, publicly available on the National Biotechnology Information Centre (NCBI) platform from five Central American countries (Belize, Guatemala, Cuba, Jamaica and Puerto Rico), were rescued from GENBANK 20 October 2020.The alignment using MEGA X (TAMURA et al., 2018)

Multiple Alignments was done using the Clustal W method with gap opening penalty of 15 and gap extension penalty of 6.6. For DNA weight matrix, we used the clustal W 1.6 with transition weight of 0.5 and divergent delay cutoff of 30%. After being aligned, the sequences were exported in Newick format, without gaps and ambiguous sites and served as INPUT for the construction of the phylogenetic tree in FIGTREE 1.4.4 software. For tree construction using FIGTREE V 1.4.4. (VLAD et al, 2008)

To assemble molecular phylogeny, we used a number of 100 pseudo-replications for Bootstrap and as an evolutionary model, we used the 2-parameter Kimura model and as rates among sites a gamma distributed with invariant sites (G+I). We used nearest-neighbor-interchange (NNI) as ML Heuristic Method, the tree generated by MEGA X, was exported in Newick format and served as input for FIGTREE 1.4.4 software. In assembling the tree in FIGTREE, we assumed the need to display the subsequent probabilities of each of the clades present, as well as the estimation of the age of each node. We also designed a time scale axis for evolutionary history and for proper sizing of this

[*]**Corresponding author:** Pierre Felix, Laboratory of Population Genetics and Computational Evolutionary Biology - LaBECom, UNIVISA, Vitoria de Santo Antao, Pernambuco, Brazil; Tel:5581984948616; E-mail: pierre.felix@hotmail.com

Received date: Oct 24, 2020; **Accepted date:** August 31, 2021; **Published date:** September 14, 2021

Citation: Felix P (2021) Phylogenetic grouping and levels of genetic structuration in SARS-CoV-2, using FIGTREE v1.4.4 software. Biol Syst Open Access. 10:5.p175

time, we defined a reverse scale on the time axis. We chose to draw the tree with thick lines and color the clades by selecting the branches. Finally, we export the tree still in NEXUS and generate graphic files in PNG.

RESULTS

General properties of analysed sequences

Of the 38 sequences analysed, with sizes ranging from 29,570 to 29,882 bp (Table 1), only 79 sites remained preserved, revealing a high degree of polymorphism for the whole set.

Table 1: General properties of the 38 complete SARS-CoV-2 genomes from five Central American Countries.

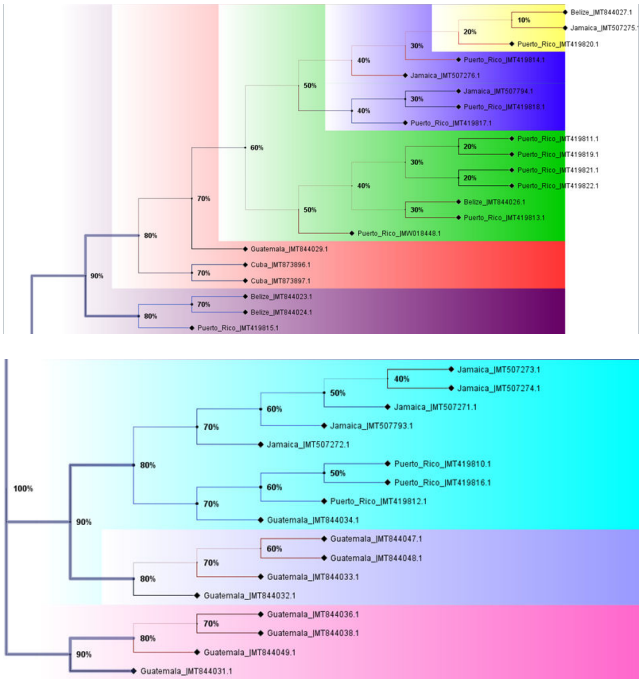
Accession	Release Date	Specie	Length	Geo Location	Host	Isolation Source	Collection Date
MT844023	05-08-2020	SARS-CoV-2	29882	Belize	Homo sapiens	-	20-03-2020
MT844024	05-08-2020	SARS-CoV-2	29882	Belize	Homo sapiens	-	01-03-2020
MT844026	05-08-2020	SARS-CoV-2	29873	Belize	Homo sapiens	-	27-03-2020
MT844027	05-08-2020	SARS-CoV-2	29882	Belize	Homo sapiens	-	09-04-2020
MT873896	11-08-2020	SARS-CoV-2	29811	Cuba	Homo sapiens	-	19-03-2020
MT873897	11-08-2020	SARS-CoV-2	29800	Cuba	Homo sapiens	-	24-03-2020
MT844029	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	15-03-2020
MT844031	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	15-03-2020
MT844032	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	17-03-2020
MT844033	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	18-03-2020
MT844034	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	19-03-2020
MT844036	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	20-03-2020
MT844038	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	21-03-2020
MT844047	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	13-03-2020
MT844048	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	15-03-2020
MT844049	05-08-2020	SARS-CoV-2	29882	Guatemala	Homo sapiens	oronasopharynx	15-03-2020
MT507271	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	09-03-2020
MT507272	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	11-03-2020
MT507273	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	14-03-2020
MT507274	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	16-03-2020
MT507275	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	16-03-2020
MT507276	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	17-03-2020
MT507793	22-05-2020	SARS-CoV-2	29885	Jamaica	Homo sapiens	swab	11-03-2020
MT507794	22-05-2020	SARS-CoV-2	29882	Jamaica	Homo sapiens	swab	16-03-2020
MT419810	01-05-2020	SARS-CoV-2	29758	Puerto Rico	Homo sapiens	oronasopharynx	23-03-2020
MT419811	01-05-2020	SARS-CoV-2	29743	Puerto Rico	Homo sapiens	oronasopharynx	23-03-2020
MT419812	01-05-2020	SARS-CoV-2	29758	Puerto Rico	Homo sapiens	oronasopharynx	23-03-2020
MT419813	01-05-2020	SARS-CoV-2	29570	Puerto Rico	Homo sapiens	oronasopharynx	23-03-2020
MT419814	01-05-2020	SARS-CoV-2	29743	Puerto Rico	Homo sapiens	oronasopharynx	23-03-2020
MT419815	01-05-2020	SARS-CoV-2	29743	Puerto Rico	Homo sapiens	oronasopharynx	24-03-2020

MT419816	01-05-2020	SARS-CoV-2	29760	Puerto Rico	Homo sapiens	oronasophary nx	24-03-2020
MT419817	01-05-2020	SARS-CoV-2	29570	Puerto Rico	Homo sapiens	oronasophary nx	23-03-2020
MT419818	01-05-2020	SARS-CoV-2	29743	Puerto Rico	Homo sapiens	oronasophary nx	23-03-2020
MT419819	01-05-2020	SARS-CoV-2	29743	Puerto Rico	Homo sapiens	oronasophary nx	24-03-2020
MT419820	01-05-2020	SARS-CoV-2	29706	Puerto Rico	Homo sapiens	oronasophary nx	24-03-2020

The analyses confirmed the presence of high and statistically significant variations, evidencing a high genetic dissimilarity among all haplotypes. However, the use of the divergence matrix in the construction of the tree helped in the recognition of minimal similarities between some haplotypes, including at different geographical points.

The maximum divergence patterns were also obtained when less robust methods of phylogenetic pairing (e.g. UPGMA) were used, reflecting the non-haplotypic structure in the clades. With the use of the divergence matrix, it was possible to identify geographical variants that had less genetic distance and a posteriori probability were able to separate the main clusters into additional small groups, confirming the presence of a minimum probability of kinship between haplotypes (Figure 1)

Figure 1: MAP Phylogenetic Tree Based.



The 38 complete SARS-CoV-2 genomes from five Central American countries (Belize, Guatemala, Cuba, Jamaica and Puerto Rico). The major geographic groupings of SARS-CoV-2 are indicated and posterior probability values are shown for key nodes. In all cases, tip times correspond to the dates (percentage) of virus sampling. Note: small clusters formed (colours) between haplotypes from different localities.

DISCUSSION

With the use of phylogenetic analysis methodologies present in MEGA X and population structuring present in software FIGTREE v 1.4.4, it was possible to detect the existence of a small degree of similarity between the haplotypes of SARS-CoV-2 of Central America.

Because significant levels of structuring were not found, we assumed that there were high levels of variation probably related to a gain of intermediate haplotypes over time, associated, perhaps, with a significant increase in gene flow. The non-occurrence of geographic isolations from past events of defragmentation may have generated a continuous pattern of genetic divergence between the groups, since the low values found for genetic distance support the presence of this continuous pattern of divergence between haplotypes, as well as in the frequency of polymorphisms. This suggests that molecular diversity may be due to synonymous substitutions as the main components of variations.

All analyses supported that the data are a phylogenetic confirmation that there is no consensus in the conservation of the SARS-CoV-2 genome in Central America, contrary to what was described by FELIX et al, 2020, for the Countries of South America, it is safe to affirm that the genetic variability of the Virus, is different in distinct subsets of countries in the American continent.

REFERENCES

1. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Molecular Biology and Evolution. 1992; 9:4 678-687.
2. Vlad I. Morariu, Balaji Vasan Srinivasan, Vikas C Raykar, Ramani Duraiswami, and Larry S. Davis. Automatic online tuning for fast Gaussian summation. Advances in Neural Information Processing Systems (NIPS), 2008.
3. Felix P T, Nascimento Filho C B, Ramos R S, Paulino A J, Venâncio D B R. Levels of genetic diversity of SARS-CoV-2 virus: reducing speculations about the genetic variability of the virus in South America. bioRxiv 2020;296491.