



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
GRADUAÇÃO EM BIOMEDICINA



MIKAELA DOS SANTOS MARINHO

EPIDEMIOLOGICAL DISTRIBUTION AND EVOLUTIONARY PROFILE OF MAYARO
VIRUS IN THE AMERICAS: AN UPDATE INTO GENOME VARIABILITY

UBERLÂNDIA

2022

MIKAELA DOS SANTOS MARINHO

EPIDEMIOLOGICAL DISTRIBUTION AND EVOLUTIONARY PROFILE OF MAYARO
VIRUS IN THE AMERICAS: AN UPDATE INTO GENOME VARIABILITY

Trabalho de Conclusão de Curso
apresentado como requisito para
obtenção do título de Bacharel em
Biomedicina na Universidade Federal
de Uberlândia.

Orientadora: Prof^ª Dr^ª Ana Carolina
Gomes Jardim

Co-orientadores: Me. Igor de
Andrade Santos; Me. Giulia
Magalhães Ferreira

UBERLÂNDIA

2022

EPIDEMIOLOGICAL DISTRIBUTION AND EVOLUTIONARY PROFILE OF
MAYARO VIRUS IN THE AMERICAS: AN UPDATE INTO GENOME
VARIABILITY

Trabalho de Conclusão de Curso
apresentado como requisito para
obtenção do título de Bacharel
em Biomedicina na
Universidade Federal de
Uberlândia.

Uberlândia, 10 de Junho de 2022.

Banca examinadora formada por:

Prof^a Dr^a Ana Carolina Gomes Jardim

Dr^a Larissa Catharina Costa

Dr. William Marciel de Souza

SUMÁRIO

ABSTRACT	7
INTRODUCTION	8
MATERIAL AND METHODS	9
Collection and alignment of MAYV genome sequences	9
Analysis of amino acid substitutions and selection pressure	9
Phylogenetic analysis of MAYV genomic sequences	10
Detection of individual recombination events	10
RESULTS.....	11
Epidemiological features of MAYV in the Americas	11
Follow-up of the amino acid substitutions and selection pressure among MAYV sequences	11
Phylogenetic relationship among MAYV complete genomic sequences	12
Individual recombination events detected among MAYV sequences	13
DISCUSSION.....	14
CONCLUSION	17
TABLES	23
CAPTIONS FOR FIGURES	26
FIGURES	28

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, por sempre acreditarem, estimularem e darem suporte aos meus sonhos

Agradeço meus amigos, por também me darem todo suporte possível, sempre me estimulando a tentar alcançar voos cada vez mais altos

Agradeço a todos meus colegas do Laboratório de Pesquisa em Antivirais, que me acolheram de braços abertos, além de repassar seus conhecimentos com muita paciência

E agradeço também aos meus orientadores, Ana Carolina, Igor e Giulia, que além de acreditarem no meu potencial, me orientaram com todo carinho do mundo.

**EPIDEMIOLOGICAL DISTRIBUTION AND EVOLUTIONARY PROFILE OF
MAYARO VIRUS IN THE AMERICAS: AN UPDATE INTO GENOME
VARIABILITY**

Mikaela dos Santos Marinho¹; Giulia Magalhães Ferreira¹; Igor Andrade Santos¹;
Victoria Riquena Grosche^{1,2}, Ana Carolina Gomes Jardim^{1,2,*}.

¹Institute of Biomedical Sciences, Federal University of Uberlândia, Uberlândia-MG
38405-302, Brazil.

²Institute of Biosciences, Humanities and Exact Sciences (Ibilce), São Paulo State
University (Unesp), Campus São José do Rio Preto, São José do Rio Preto, SP, Brazil

***Corresponding Author:** Professor Ana Carolina Gomes Jardim, Institute of
Biomedical Sciences, ICBIM, Federal University of Uberlândia, Avenida Amazonas, 4C-
Room 216, Umuarama, Uberlândia, Minas Gerais, Brazil, CEP: 38405-302

Tel: +55 (34) 3225-8682

E-mail: jardim@ufu.br

ABSTRACT

Mayaro virus (MAYV) is the etiologic agent of Mayaro Fever, a disease characterized by dengue-like symptoms and arthralgias that can progress to a chronic condition. MAYV is phylogenetically classified into the genotypes D (widely dispersed), L (limited), and N (new), being the D the most common. Additionally, a new genotype was generated by a recombination event between L and D, resulting in the genotype L/D. There are limited epidemiological and evolutionary studies on MAYV, resulting in the difficulty in tracking the emerging potential of this virus. In this context, here we investigated the MAYV genomic variations in sequences from different regions of America, and characterized the phylogenetic relationship among virus strains. Through the analysis of 40 complete genome sequences of MAYV from the GenBank database, uploaded in South, Central, and North American Countries, the phylogenetic tree was reconstructed, showing that sequences grouped according to the genotypes L, D, and L/D. Sequences characterized as genotype D were closely related to sequences from their respective countries, suggesting that isolates may have originated from circulating lineages in the respective countries. In addition, an unidentified sequence from the United States of America grouped with the genotype D, suggesting the insertion of this genotype in the country. Furthermore, our data demonstrated that a sequence previously characterized as genotype L grouped with isolates of genotype L/D, and presented an insertion of nucleotides into the nsP3, positioned identically to the genotype L/D. These results were also corroborated by our recombination analysis, which also identified this isolate as L/D hybrid. The analysis of amino acid mutations among sequences indicated positive and negative selective pressure sites, suggesting adaptability of viral fitness. These results provide information on the epidemiological distribution and evolutionary profile of MAYV, contributing to a better understanding of genotypes circulating in the Americas, as well as the potential viral changes that may contribute to the emergence of mutation of interest or new viral genotypes.

Keywords: Arboviruses, bioinformatics, genomic variability, Mayaro virus, Mayaro fever.

INTRODUCTION

The Mayaro virus (MAYV) was first identified in 1954 in Trinidad and Tobago, in samples of febrile forest workers (Diagne et al., 2020), being later described as the etiologic agent of Mayaro Fever. This disease is characterized by fever, headache, vomit, diarrhea, muscle pain, skin rashes, and long-lasting arthralgia (Caicedo et al., 2021). In severe cases, MAYV infection can result in chronic polyarthrititis, neurological complications, hemorrhage, and myocarditis, which can lead to death (Acosta-Ampudia et al., 2018).

Mayaro fever in humans is mainly observed in Central and South Americas, mainly in the Amazon Forest and nearby regions, however, it was also reported in subtropical regions (Caicedo et al., 2021). Additionally, the dengue-like symptoms, associated with the endemic areas common to other arboviruses, poses as a challenge to differential diagnosis and contribute to the underreporting of Mayaro Fever cases and genome sequences available on databanks. What is more, there is no data about the circulating virus in other tropical regions, such as Africa and Asia (Diagne et al., 2020).

The main vectors for MAYV transmission are *Haemagogus sp* mosquitoes (Diagne et al., 2020). However, the virus was also isolated from different nonurban arthropod genera such as *Culex*, *Mansonia*, *Psorophora*, *Sabethes*, *Aedes albopictus*, and *Aedes aegypti*, mosquitoes which present high susceptibility to MAYV transmission, indicating intense viral adaptability to new vectors (Ganjian and Riviere-Cinamond, 2020). The distribution of these species in tropical and subtropical countries associated with globalization is an aggravating factor, increasing the risk of MAYV dissemination (Diagne et al., 2020).

The MAYV is classified in the *Togaviridae* family, *Alphavirus* genus. The virion possesses a positive single-stranded RNA of approximately 11.5 kb, associated with a capsid protein (C), and involved by a host-lipid membrane envelope inserted of the E1 and E2 glycoproteins (Choi et al., 2019). The MAYV genome is constituted of two open reading frames (ORF), one in the 5' UTR region, responsible for encoding the non-structural proteins (nsP1, nsP2, nsP3, and nsP4), and the second after the nsP4 gene, which encodes the structural proteins (CP, E3, E2, 6k, E1) (Acosta-Ampudia et al., 2018).

MAYV is classified in the genotypes D (widely dispersed), L (limited), and N (new), being the latest represented by a single sequence isolated in Peru (Diop et al.,

2019). It is also important to emphasize the higher rates of mutations within RNA viruses, which leads to an intra-host variability of the viral population (Duffy, 2018). Therefore, phylogenetic analyses represent important methodologies for developing epidemiological studies, allowing the correlation of data derived from different regions and presuming the history of the dissemination of viral variants (Chesnut et al., 2019). Currently, only a few epidemiological studies on MAYV are reported, with a limited number of genome sequences available on databanks. Considering the severe symptoms caused by MAYV, and that this virus is responsible for a neglecting disease, allied to the possibility of the emergence of new strains, or additional transmitting vectors, the study of the viral genome variability represents an essential tool to investigate the virulence and transmissibility of the different viral genotypes (Chesnut et al., 2019).

In this context, genetic and phylogenetic analyses of MAYV complete genome sequences were carried out here to investigate genomic variations and the phylogenetic relationships among the MAYV strains from samples collected in the Americas. Thus, an overview of the main features of the MAYV genomic variability was obtained, indicating evolutionary forces that could drive the evolution of the viral genome, and possibly contribute to viral adaptation.

MATERIAL AND METHODS

Collection and alignment of MAYV genome sequences

Sequences of the complete genome of MAYV (about 11,147 nucleotides) were collected from the GenBank database, and organized according to the year of collection, origin, genotype, host, length of the sequenced genetic material, article of reference, and reference of the genotype identification (**Supplementary Table 1**). The dataset containing all the sequences was submitted to alignment by the ClustalW multiple alignment method, using the program Bioedit 7.2 (Sagnelli et al., 2020). The sequence DQ001069.1 was used as a reference sequence, and the parameters of alignment used were defined by the program default configuration.

Analysis of amino acid substitutions and selection pressure

Sequences were aligned according to their genotype, resulting in two datasets: 1) sequences identified as genotype D, using DQ001069.1 as the reference sequence; 2) sequences identified as genotypes L and L/D, employing as reference the genotype L sequence NC_003417.1. The alignments were edited according to the protein-coding regions and the translation was performed. Amino acid substitutions, in comparison to the reference sequence, were analyzed. The selective pressure analysis was performed on the above reported MAYV protein-coding sequence sub-sets through the Datamonkey Adaptive Evolution Server (Delpont et al., 2010), in order to identify and localize statistically supported positive and negative selective pressure sites, as previously described (Grosche et al., 2021). To predict the impact of each substitution in protein selection, the FUBAR model (Fast Unconstrained Bayesian AppRoximation) was employed, which uses a maximum-likelihood (ML) approach to infer nonsynonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny (Murrell et al., 2013). The selection pressure for each site was considered constant along the entire phylogeny. The statistically supported selective pressure sites considering Bayes factor (BF) >30 were reported. As a complementary analysis, the amino acid substitutions were also evaluated using the PROVEAN software version 1.1.3 (Choi and Chan, 2015).

Phylogenetic analysis of MAYV genomic sequences

The phylogenetic tree was reconstructed employing PhyML 3.0 through the ATGC- Montpellier Bioinformatics Platform (Guindon et al., 2010), using maximum-likelihood analysis (ML), and the GTR+G+I model of substitution, as determined by the Smart Model Selection in PhyML (Lefort et al., 2017). A thousand replicates were used to test the support given by the data to the clusters of the tree topology, and bootstrap values >70 were considered significant (McCormack & Clewley, 2002). The phylogenetic tree was edited using the program FigTree v.1 4.4 (Sagnelli et al., 2020).

Detection of individual recombination events

The 40 complete genome sequences of MAYV were evaluated for potential recombination events using Recombination Detection Program 4 (RDP4) (Martin et al., 2015). RDP4 was used with the default parameters, applying all available methods through the full exploratory recombination scan (Recombination Detection Program, Bootscan, MaxChi, GENECONV, Chimaera, SiScan, 3Seq, LARD, and Phylpro). The

beginning and the end of the breakpoints of these events suggested by RPD4 were used to identify the potential recombinant sequences.

RESULTS

Epidemiological features of MAYV in the Americas

The 40 complete genome sequences collected from GenBank were uploaded to the database, between 1954 and 2015, in Brazil, Trinidad and Tobago, Haiti, United States of America (USA), Bolivia, Peru, French Guiana, and Venezuela (**Figure 1**). Two sequences were reported in the literature as hybrids L/D, eleven as genotype L, twenty-six as genotype D, and one, collected from an imported case in the USA, presented no classification. All MAYV sequences and the related data employed in this study are shown in **Supplementary Table 1**.

Follow-up of the amino acid substitutions and selection pressure among MAYV sequences

The analysis of the amino acid substitutions and their effects on viral protein structures was performed using DataMonkey and PROVEAN programs to detect the presence of neutral (synonymous), adaptative (non-synonymous), and purifying (non-synonymous) substitutions, under a P-value < 0.05 (**Figure 2**). Additionally, the Bayes Factor (BF) was also employed to measure the relative evidence of substitutions, being values higher than 30 characterized as very strong evidence of the positive selection pressure hypothesis at the site (Dittrich et al., 2019). As a result, several substitutions were identified among the MAYV sequences (**Supplementary Table 2**). Among them, 6 sites were identified under positive/adaptative selection (three in genotype D and three in genotype L – L/D) (**Table 1; Figure 2**). For the genotype D, the amino acid substitutions S487G and V386M were observed in nsP1, with Bayes Factor (BF) of 79.944 and 54.772, respectively (**Table 1; Figure 2**). Additionally, the amino acid change G586V was identified in the nsP4, with a BF of 62.506. On the other hand, the genotypes L and L/D demonstrated amino acid substitutions in E2 (R110V: BF of 31.703; F381V: BF of 30.265) and E1 (T300L: BF of 39.116) proteins (**Table 1; Figure 2**).

The sequences of the MAYV genome also presented amino acid substitutions characterized as neutral or under negative selection pressure in all viral proteins

(**Supplementary Tables 2 - 4**). Among them, only those with a frequency of over 60% are presented in **Tables 2 and 3**. The complete data on substitution frequency is available in **Supplementary Table 2**. The most frequent amino acid substitutions were observed in the majority of the MAYV proteins, mainly in sequences of the genotypes L and L/D (**Table 2**). The insertion of the amino acids PACQDT (nucleotide sequence: AGCGTGCCAAGACAC) into the position 443-448 of the nsP3 was found in three sequences, two of them already described as L/D (KX496990.1 - Haiti, 2015; KT818520.1 - Brazil, 2014), and one as L (KY985361.1 - Haiti, 2014) (**Figure 3**). Alternatively, sequences of the MAYV genotype D demonstrated a lower rate of amino acid substitutions than sequences of genotypes L and L/D, with most of them concentrated into the non-structural proteins nsP2 and nsP3 (**Table 3**).

Phylogenetic relationship among MAYV complete genomic sequences

The phylogenetic analysis revealed that sequences grouped according to the MAYV genotypes (D, L, and L/D) (**Figure 4**). Additionally, sequences characterized as genotype D isolated in Bolivia (1955 and 2002), Peru (1955, 1997, and 2000), and Brazil (1982, 1991, 2008, 2009, 2011 and 2012), were closely related to sequences from their respective countries, sustained by a bootstrap over 80% (**Figure 4**). It suggests that the clustered sequences may have originated from circulating lineages in the respective countries due to the phylogenetic relationship among the isolates. Furthermore, the sequence MT227562.1, isolated from a sample of an allochthonous case in Louisiana (USA), which had no previous description regarding its genotype, formed a cluster with sequences of the genotype D (**Figure 4**), suggesting this sequence as belonging to this genotype. This strain was grouped with isolates MK573240.1 (1957) and MK070492.1 (1984) from Trinidad and Tobago, supported by a bootstrap of 100%.

The topology of the tree demonstrated that L/D sequences formed a cluster with sequences of the L genotype, suggesting higher genomic similarity (bootstrap of 100%). The sequences KY985361.1 - Haiti, 2014 (previously reported as genotype L) (Caicedo et al., 2021) and KX496990.1 - Haiti, 2015 (L/D), showed phylogenetic proximity with one of the L/D sequences from Brazil (KT818520.1, 2014), supported by a bootstrap of 100% (**Figure 4**). Interestingly, these sequences were collected in subsequent years (2014 – 2015), which may suggest a possible route of transmission of MAYV. Additionally, the sequence KY985361.1 (Haiti, 2014), previously characterized as the genotype L (Caicedo et al., 2021), clustered with the L/D hybrids isolates (**Figure 4**). This sequence presented

the insertion of nucleotides AGCGTGCCAAGACACG into the nsP3 protein, positioned identically to the insertions found in the L/D genotype sequences KX496990.1 (Haiti, 2015) and KT818520.1 (Brasil, 2014) (**Figure 3**).

Individual recombination events detected among MAYV sequences

Due to the genetic and phylogenetic similarities between the sequence KY985361.1 (Haiti, 2014), previously classified as genotype L, and the L/D hybrids isolates KX496990.1 (Haiti, 2015) and KT818520.1 (Brazil, 2014), a recombination analysis was performed using the RDP4 tool (MARTIN, et al, 2000). The program applies several methods to detect recombination events, including RDP4 (Recombination Detection Program), Bootscan, MaxChi, GENECONV, Chimaera, SiScan, and 3Seq. By submitting the complete alignment with all analyzed sequences of genotypes D, L, and L/D, RDP4 identified evidence of recombination exactly among the sequences KY985361.1 (Haiti, 2014), KX496990.1 (Haiti, 2015), and KT818520.1 (Brazil, 2014) (**Figure 5**). The recombination event was positive in all seven methods listed above ($p < 0.05$), being inconclusive only on LARD and Phylpro, also used in the analysis of the sequences (**Supplementary Table 5**).

The sequence KX496990.1 (Haiti, 2015) was identified as a hybrid between KM400591.1 (genotype D, Brazil, 2004) and KY618133.1 (genotype L, Brazil, 1988) (**Figure 5**). The L genotype sequence was the most closely related to the likely L/D hybrid, being a major parent, and the D genotype sequence was the minor parent. Furthermore, the sequence KT818520.1 (Brasil, 2014) was identified as an L/D hybrid, with a probability of having originated from the same recombination event as KX496990.1 (Haiti, 2015) (**Figure 5**). Additionally, KY985361.1 (Haiti, 2014), previously reported as genotype L (Caicedo et al., 2021), was also identified here as an L/D hybrid by all RDP4 conclusive analysis. The sequence was identified as possibly originating from the major parent KY618133.1 (Genotype L, Brazil, 1988) and the minor parent KY618131.1 (Genotype D, Brazil, 1978) (**Figure 5**). Interestingly, as the other sequences, KY985361.1 (Haiti, 2014) also showed higher similarity to its major parent of genotype L. All the recombination events mentioned presented p -values $\leq 4.8 \times 10^{-09}$ (**Supplementary Table 5**), being considered potential recombinant events.

DISCUSSION

The incidence of the MAYV infections has been mainly described in Latin America, with most Mayaro fever cases reported in Brazil (Ganjian and Riviere-Cinnamond, 2020). In agreement with this data, the majority of genomic sequences analyzed here were obtained from samples collected in Brazil, which included MAYV isolates of the genotypes L, D, and L/D. Although most cases of Mayaro fever have been observed in South America, there are reports of outbreaks in Haiti, indicating that the virus is also circulating in Central America (Blohm et al., 2019). Mayaro Fever was likewise reported in Europe (Acosta-Ampudia et al., 2018), however, there are no epidemiological studies that describe the introduction of the genotypes in the region, or whether the infection has spread by local vectors. Furthermore, to the best of our knowledge, there is no current data on the circulation of MAYV in Africa and Asia, due to the lack of data reported by these continents (Diagne et al., 2020). It is also observed that there is poorly described information concerning the MAYV strains, such as the N genotype, isolated in Peru in 2010, with no current description of its dissemination (Caicedo et al., 2021). Therefore, studies that aim to investigate the epidemiological characteristics of MAYV worldwide, as well as the genetic and phylogenetic relationship among circulating viral strains are essential.

Here we analyzed the MAYV genomic variations in sequences from different regions of America and characterized the phylogenetic relationship among these virus strains, in order to investigate the main evolutionary forces driving the evolution of the viral genome and their possible effect on viral activity, associating them to the dissemination of MAYV. Concerning the genetic variability among the sequences of the complete genome of MAYV, it is important to emphasize the higher rates of mutations within RNA viruses, which leads to an intra-host variability of the viral population, as well as the emergence of new species, genera, and/or new viral families (Duffy, 2018). Herein, a high incidence of mutations was found in the analyzed sequences, with several of those substitutions classified as purifying mutations, which demonstrates mechanisms to eliminate deleterious mutations and maintain the accumulated characteristics of the virus (Lin et al., 2019). These mutations were widely observed when the genomic regions related to most of viral proteins were evaluated, some of them directly involved in viral replication. Through the analysis carried out here, substitutions in the E2 protein were identified, such as the Q247R on genotype D, and R245A and M246E on genotype L-L/D, all classified as neutral in the selection pressure analysis. The E2 glycoprotein is

essential for the interaction of the virus with the host cell receptor, and to mediate the fusion through the E1 protein and the endosome after virus entry into the host cell (Choi et al., 2019). Previous studies showed that viruses containing a lysine to arginine change at residue 253 of Getah virus E2 protein increased infectivity in mammalian cells and reduced virulence in a mouse model (Wang et al., 2022), and therefore, may suggest that the region between the residues 240 and 260 is currently under selection, being the most exposed region of the E2 glycoprotein trimer, hence is under the environmental and physical forces (Choi et al., 2019).

What is more, the presence of positive selection pressure sites in sequences of the genotype L was observed only in the sequence MK573239.1 (Brazil, 1984) at E1 and E2 proteins. Interestingly, the mutations into E1 (T300L) and E2 (R110V and F381V) proteins were not found in other isolates from samples collected in the next years, which may suggest that these substitutions did not sustained their circulation in low frequency. Earnest and colleagues evaluated different monoclonal IgG antibodies against the MAYV strain containing the same mutations on E2 protein under positive selection described here (R110V and F381V). As an outcome, only 4 of 18 antibodies demonstrated elite neutralizing activity against this strain of MAYV, which emphasizes the role of the mutation on escaping immune response and potential increased infectivity, even though it was not detected in other sequences (Earnest et al., 2019). In addition, mutations identified in the genomic regions of the nsP1 and nsP4 of sequences of the genotype D were found to be under positive selection pressure (**Figure 2, Supplementary Table 2**), suggesting possible viral adaptability and improvement of fitness during viral replication. The nsP1 protein is directly involved in the initiation of virus replication, being responsible for interacting with the inner face of the plasma membrane and promoting the formation of the spherules, which are vesicles where the replicative process initiates (Zhang et al., 2021). The nsP4 is an RNA-dependent RNA polymerase (RdRp) responsible for the production of new viral RNAs (Lello et al., 2021), required for alphavirus infection, generating positive-strand genomic and subgenomic (sg) RNAs (Lello et al., 2021). Even though the presence of mutations by positive selective pressure is relevant due to its impact on virus evolution, the low level of positive selection suggests that the amino acid changes have been a result of random genetic drift. Interestingly, the same pattern was seen in other RNA viruses such as the West Nile virus and Hepatitis C virus (Beasley et al., 2003).

Alternatively, negative selective pressure was observed in several positions of the viral genome, which demonstrates an attempt to constrain the variability at the protein level, since the negative selection tends to conserve the sequence and the structural characteristics of active motifs of the proteins (Domingo et al., 1998). The proteins from the replicative complex (nsP1-4), as well as the structural proteins E2 and E1, were found under negative pressure in several positions in almost all analyzed sequences, which may suggest that the circulating genotypes are with an adapted fitness. The genetic variability and the evolution potential within a population are the product of a balance of selection and genetic drift (Domingo et al., 1998), and therefore, further analysis regarding the up and down fluctuation of viral fitness is necessary to achieve knowledge of the gain or loss of function due to substitutions observed over the viral genome. However, the restricted number of sequences of MAYV poses a delay in the analyses of the persistence and disappearance of mutations over time, as well as difficult the suggestion of adaptability to the host, higher transmissibility, or immune escape (Cong et al., 2007)

Our phylogenetic analysis demonstrated that the sequence MT227562.1 was classified as genotype D. This sequence was obtained from a sample of an allochthonous case in Louisiana (USA), isolated from *Icterus spurius*, a bird with a migratory route between Central America, Colombia/Venezuela, and the USA (DeBrock et al., 2021). Therefore, it is possible to infer that the bird was infected in the northern South or Central America (similarity with sequences from Trinidad and Tobago) and, during its migration to the USA, led to the insertion of the genotype in the region, given the predominance of genotype D in the Americas. Additionally, the L/D sequences were phylogenetically closer to the sequences of genotype L, demonstrating higher genomic similarity with the L genotype sequences, as indicated by previous studies (Mavian et al., 2017). It is important to emphasize that L/D hybrids were initially classified as members of the L genotype, prior to the recognition of the L/D recombination (Lednicky et al., 2016; Mota et al., 2015). Sequences of the genotype L/D (KY985361.1, 2014 and KX496990.1, 2015) from Haiti clustered with the L/D sequence from Brazil (KT818520.1, 2014), with one year of difference between them. Among these, KT818520.1 (Brazil, 2014) and KX496990.1 (Haiti, 2015) were identified by RDP4 as potentially originating from the same recombination event, which may suggest a possible route of dissemination of MAYV. Furthermore, the emergence of MAYV recombinants/hybrids demonstrates the potential for the emergence of a future new genotype, and its dissemination, highlighting

the need for genomic surveillance for this pathogen. It has also been described for other RNA viruses, such as in Chikungunya virus (CHIKV), that through homologous recombination resulted in cross-species transmission, and therefore, impacted virus evolution, emergence, and epidemiology (Casal et al., 2015). In addition, the interspecies recombination was also identified in other Alphavirus, such as Eastern equine encephalitis virus (EEEV) and Sindbis-like virus (SIN), originating from the Western equine encephalitis virus (WEEV), Highlands J virus (HJV), and Fort Morgan virus (FMV) (Allison et al., 2015). In this context, MAYV recombinants could emerge and spread as new strains with increased fitness in human and *Haemagogus* vector hosts, as well as adapt into new hosts.

CONCLUSION

In conclusion, the data presented here demonstrate that sequences of the complete genome of MAYV are under negative and positive selection pressures, emphasized by the high number of purifying selection sites, that might result in the emergence of new strains. Additionally, it shows the importance of diagnosing and sequencing the MAYV genome to characterize virus adaptation and transmissibility. Therefore, the virus surveillance demonstrates to be essential to define its degree of dissemination and implement protective measures to manage Mayaro fever.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any condition that could be a potential conflict of interest.

Authors and contributors

Collation of data and results analysis: M. S. M., I. A. S., G. M. F.; Drafting of the manuscript: M. S. M., I. A. S., G. M. F.; Figures: V.R.G.; Study design, supervision, and critical revision: I. A. S., G. M. F., and A. C. G. J. All authors reviewed the manuscript.

Acknowledgments

ACGJ is grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Brasil—Prevention and Combat of Outbreaks, Endemics, Epidemics and Pandemics—Finance Code #88881.506794/2020-01, and to FAPEMIG (Minas Gerais Research Foundation APQ-03385-18). IAS thanks to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for the scholarship # 142495/2020-4. GMF and VRG thanks to CAPES for the scholarship # 88887.571465./2020-00 and # 88887.505971/2020-00.

REFERENCES

- Acosta-Ampudia, Y., Monsalve, D.M., Rodríguez, Y., Pacheco, Y., Anaya, J.-M., Ramírez-Santana, C., 2018. Mayaro: an emerging viral threat? *Emerging Microbes & Infections* 7, 1–11. <https://doi.org/10.1038/s41426-018-0163-5>
- Allison, A.B., Stallknecht, D.E., Holmes, E.C., 2015. Evolutionary genetics and vector adaptation of recombinant viruses of the western equine encephalitis antigenic complex provides new insights into alphavirus diversity and host switching. *Virology* 474, 154–162. <https://doi.org/10.1016/j.virol.2014.10.024>
- Beasley, D.W.C., Davis, C.T., Guzman, H., Vanlandingham, D.L., Travassos da Rosa, A.P.A., Parsons, R.E., Higgs, S., Tesh, R.B., Barrett, A.D.T., 2003. Limited evolution of West Nile virus has occurred during its southwesterly spread in the United States. *Virology* 309, 190–195. [https://doi.org/10.1016/S0042-6822\(03\)00150-8](https://doi.org/10.1016/S0042-6822(03)00150-8)
- Blohm, G., Elbadry, M.A., Mavian, C., Stephenson, C., Loeb, J., White, S., Telisma, T., Chavannes, S., Beau De Rochar, V.M., Salemi, M., Lednicky, J.A., Morris, J.G., 2019. Mayaro as a Caribbean traveler: Evidence for multiple introductions and transmission of the virus into Haiti. *International Journal of Infectious Diseases* 87, 151–153. <https://doi.org/10.1016/j.ijid.2019.07.031>
- Caicedo, E.-Y., Charniga, K., Rueda, A., Dorigatti, I., Mendez, Y., Hamlet, A., Carrera, J.-P., Cucunubá, Z.M., 2021. The epidemiology of Mayaro virus in the Americas: A systematic review and key parameter estimates for outbreak modelling. *PLOS Neglected Tropical Diseases* 15, e0009418. <https://doi.org/10.1371/journal.pntd.0009418>

- Casal, P.E., Chouhy, D., Bolatti, E.M., Perez, G.R., Stella, E.J., Giri, A.A., 2015. Evidence for homologous recombination in Chikungunya Virus. *Molecular Phylogenetics and Evolution* 85, 68–75. <https://doi.org/10.1016/j.ympev.2015.01.016>
- Chesnut, M., Muñoz, L.S., Harris, G., Freeman, D., Gama, L., Pardo, C.A., Pamies, D., 2019. In vitro and in silico Models to Study Mosquito-Borne Flavivirus Neuropathogenesis, Prevention, and Treatment. *Frontiers in Cellular and Infection Microbiology* 9. <https://doi.org/10.3389/fcimb.2019.00223>
- Choi, H., Kudchodkar, S.B., Reuschel, E.L., Asija, K., Borole, P., Ho, M., Wojtak, K., Reed, C., Ramos, S., Bopp, N.E., Aguilar, P. v., Weaver, S.C., Kim, J.J., Humeau, L., Tebas, P., Weiner, D.B., Muthumani, K., 2019. Protective immunity by an engineered DNA vaccine for Mayaro virus. *PLOS Neglected Tropical Diseases* 13, e0007042. <https://doi.org/10.1371/journal.pntd.0007042>
- Choi, Y., Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31, 2745–2747. <https://doi.org/10.1093/bioinformatics/btv195>
- Cong, M., Heneine, W., García-Lerma, J.G., 2007. The Fitness Cost of Mutations Associated with Human Immunodeficiency Virus Type 1 Drug Resistance Is Modulated by Mutational Interactions. *Journal of Virology* 81, 3037–3041. <https://doi.org/10.1128/JVI.02712-06>
- DeBrock, S., Cohen, E., Balasubramanian, S., Marra, P.P., Hamer, S.A., 2021. Characterization of the Plasmodium and Haemoproteus parasite community in temperate-tropical birds during spring migration. *International Journal for Parasitology: Parasites and Wildlife* 15, 12–21. <https://doi.org/10.1016/j.ijppaw.2021.03.013>
- Delpont, W., Poon, A.F.Y., Frost, S.D.W., Kosakovsky Pond, S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26, 2455–2457. <https://doi.org/10.1093/bioinformatics/btq429>
- Diagne, C.T., Bengue, M., Choumet, V., Hamel, R., Pompon, J., Missé, D., 2020. Mayaro Virus Pathogenesis and Transmission Mechanisms. *Pathogens* 9, 738. <https://doi.org/10.3390/pathogens9090738>

- Diop, Alout, Diagne, Bengue, Baronti, Hamel, Talignani, Liegeois, Pompon, Morales Vargas, Nougairède, Missé, 2019. Differential Susceptibility and Innate Immune Response of *Aedes aegypti* and *Aedes albopictus* to the Haitian Strain of the Mayaro Virus. *Viruses* 11, 924. <https://doi.org/10.3390/v11100924>
- Dittrich, D., Leenders, R.Th.A.J., Mulder, J., 2019. Network Autocorrelation Modeling: A Bayes Factor Approach for Testing (Multiple) Precise and Interval Hypotheses. *Sociological Methods & Research* 48, 642–676. <https://doi.org/10.1177/0049124117729712>
- Domingo, E., Escarmís, C., Sevilla, N., Baranowski, E., 1998. Population Dynamics in the Evolution of RNA Viruses. pp. 721–727. https://doi.org/10.1007/978-1-4615-5331-1_93
- Duffy, S., 2018. Why are RNA virus mutation rates so damn high? *PLOS Biology* 16, e3000003. <https://doi.org/10.1371/journal.pbio.3000003>
- Earnest, J.T., Basore, K., Roy, V., Bailey, A.L., Wang, D., Alter, G., Fremont, D.H., Diamond, M.S., 2019. Neutralizing antibodies against Mayaro virus require Fc effector functions for protective activity. *Journal of Experimental Medicine* 216, 2282–2301. <https://doi.org/10.1084/jem.20190736>
- Ganjan, N., Riviere-Cinamond, A., 2020. Mayaro virus in Latin America and the Caribbean. *Revista Panamericana de Salud Pública* 44, 1. <https://doi.org/10.26633/RPSP.2020.14>
- Grosche, V.R., Santos, I.A., Ferreira, G.M., Dutra, J.V.R., Costa, L.C., Nicolau-Junior, N., Queiroz, A.T.L., José, D.P., Jardim, A.C.G., 2021. Insights on the SARS-CoV-2 genome variability: the lesson learned in Brazil and its impacts on the future of pandemics. *Microbial Genomics* 7. <https://doi.org/10.1099/mgen.0.000656>
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59, 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Lednický, J., de Rochars, V.M.B., Elbadry, M., Loeb, J., Telisma, T., Chavannes, S., Anilis, G., Cella, E., Ciccozzi, M., Okech, B., Salemi, M., Morris, J.G., 2016.

- Mayaro Virus in Child with Acute Febrile Illness, Haiti, 2015. *Emerging Infectious Diseases* 22, 2000–2002. <https://doi.org/10.3201/eid2211.161015>
- Lefort, V., Longueville, J.-E., Gascuel, O., 2017. SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution* 34, 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Lello, L.S., Bartholomeeusen, K., Wang, S., Coppens, S., Fragkoudis, R., Alphey, L., Ariën, K.K., Merits, A., Utt, A., 2021. nsP4 Is a Major Determinant of Alphavirus Replicase Activity and Template Selectivity. *Journal of Virology* 95. <https://doi.org/10.1128/JVI.00355-21>
- Lin, J.-J., Bhattacharjee, M.J., Yu, C.-P., Tseng, Y.Y., Li, W.-H., 2019. Many human RNA viruses show extraordinarily stringent selective constraints on protein evolution. *Proceedings of the National Academy of Sciences* 116, 19009–19018. <https://doi.org/10.1073/pnas.1907626116>
- Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1. <https://doi.org/10.1093/ve/vev003>
- Mavian, C., Rife, B.D., Dollar, J.J., Cella, E., Ciccozzi, M., Prosperi, M.C.F., Lednicky, J., Morris, J.G., Capua, I., Salemi, M., 2017. Emergence of recombinant Mayaro virus strains from the Amazon basin. *Scientific Reports* 7, 8718. <https://doi.org/10.1038/s41598-017-07152-5>
- McCormack, G.P., Clewley, J.P., 2002. The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. *Reviews in Medical Virology* 12, 221–238. <https://doi.org/10.1002/rmv.355>
- Mota, M.T.O., Vedovello, D., Estofolete, C., Malossi, C.D., Araújo, J.P., Nogueira, M.L., 2015. Complete Genome Sequence of Mayaro Virus Imported from the Amazon Basin to São Paulo State, Brazil. *Genome Announcements* 3. <https://doi.org/10.1128/genomeA.01341-15>
- Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., Scheffler, K., 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution* 30, 1196–1205. <https://doi.org/10.1093/molbev/mst030>

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>

Sagnelli, C., Uberti-Foppa, C., Bagaglio, S., Cella, E., Scolamacchia, V., Hasson, H., Salpietro, S., Messina, E., Morsica, G., Angeletti, S., Ciccozzi, M., Lazzarin, A., Sagnelli, E., 2020. Molecular epidemiology of HIV-1 infection in immigrant population in northern Italy. *Epidemiology and Infection* 148, e19. <https://doi.org/10.1017/S0950268819002012>

Wang, N., Zhai, X., Li, X., Wang, Y., He, W.-T., Jiang, Z., Veit, M., Su, S., 2022. Attenuation of Getah Virus by a Single Amino Acid Substitution at Residue 253 of the E2 Protein that Might Be Part of a New Heparan Sulfate Binding Site on Alphaviruses. *Journal of Virology* 96. <https://doi.org/10.1128/jvi.01751-21>

Zhang, K., Law, Y.-S., Law, M.C.Y., Tan, Y.B., Wirawan, M., Luo, D., 2021. Structural insights into viral RNA capping and plasma membrane targeting by Chikungunya virus nonstructural protein 1. *Cell Host & Microbe* 29, 757-764.e3. <https://doi.org/10.1016/j.chom.2021.02.018>

TABLES

Table 1. Sites of the MAYV genome under positive selection pressure based on FUBAR that presented amino acid substitutions.

Genotype	Protein	Site	α	β	Probability $\alpha > \beta$	Probability $\alpha < \beta$	Bayes Factor (BF)
D	nsp1	487	0.948	17.836	0.009	0.978	79.944
		386	1.61	21.118	0.012	0.969	54.772
	nsp4	586	0.888	13.416	0.018	0.965	62.506
L L/D	E2	110	1.671	18.74	0.022	0.954	31.703
		381	1.618	17.723	0.024	0.952	30.265
	E1	300	1.722	22.149	0.018	0.961	39.116

Table 2. Amino acid substitutions characterized as neutral or under negative selection pressure with frequency over 60% in sequences of Genotype L and L/D.

Frequent Mutations on Genotype L and L/D frequent mutations		
Protein	Substitution	Rate (%)
nsP1	T234K	84.6
	V518A	69.2
nsP2	S47Q	84.6
	S127R	84.6
	L219H	84.6
	M229I	84.6
	D305H	84.6
	Q571R	84.6
	S724R	84.6
	T725A	84.6
	A727G	84.6
	T751R	84.6
	V752P	84.6
	R753P	84.6
nsP3	A3V	76.9
	E372D	53.8
	V389I	76.9
	S416T	84.6
nsP4	Y93S	84.6
	A98E	84.6
	R582E	84.6
	L583R	84.6
C	P220R	84.6
E3	T365I	61.5
	F381L	61.5
	R110A	76.9
	R303P	76.9
E2	S201C	84.6
	Q206K	84.6
	A207P	84.6
	V208S	84.6
	R245A	84.6
	M246E	84.6
	F299L	84.6
	F337L	84.6
	V338W	84.6
	V339S	84.6
	A340Q	84.6
	R372A	84.6
	T393N	84.6
6K	A32T	76.9
E1	R191P	84.6

Table 3. Amino acid substitutions characterized as neutral or under negative selection pressure with frequency over 60% in sequences of Genotype D.

Frequent Mutations on Genotype D		
Protein	Substitution	Rate (%)
nsP1	S442R	92.6
	F483S	88.8
nsP2	R49K	92.6
	V317I	77.8
	V341I	81.5
	A528V	96.3
nsP3	V168E	92.6
	R374Q	81.4
	V380A	62.9
	R385Q	81.4
	H411R	81.4
nsP4	I103V	85.2
C	-	-
E3	-	-
E2	A59T	70.3
	D344E	70.3
6K	-	-
E1	I34V	76.9
	S360A	76.9

CAPTIONS FOR FIGURES

Figure 1: Geographical distribution of the analyzed sequences collected from GenBank. The 40 complete genome sequences selected in the study were collected between 1954 and 2015, in Brazil, Trinidad and Tobago, Haiti, United States of America (USA), Bolivia, Peru, French Guiana, and Venezuela. The image was generated employing GIMP 2.10.20 (www.gimp.org).

Figure 2: Amino acid substitutions and selection pressure among MAYV sequences. The most relevant amino acid substitutions among the sequences of complete genome of MAYV that presented positive selection pressure with a Bayes factor (BF) >30 are shown (BF of 30–100: very strong evidence for positive selection pressure). The sequences analyzed were related to samples collected between 1954 and 2015, available on the GenBank database up to May 2022. DQ001069.1 isolated from sample of *Homo sapiens* in French Guiana was used as the reference sequence. The frequency of the substitutions is presented as percentage, where 0% is represented in light orange and 100% in dark brown. The image was generated employing GIMP 2.10.20 (www.gimp.org).

Figure 3: Schematic representation of the nucleotide insertion in the nsP3 observed in the sequences KY985361.1 (Haiti, 2014), KX496990.1 (Haiti, 2015), and KT818520.1 (Brazil, 2014). The image was generated employing GIMP 2.10.20 (www.gimp.org).

Figure 4: Phylogenetic tree reconstruction using 40 sequences of the MAYV complete genome. The sequences were aligned employing ClustalW multiple alignment method and the phylogenetic tree reconstructed employing PhyML 3.0, through the ATGC-Montpellier Bioinformatics Platform, using maximum-likelihood analysis (ML), and the GTR+G+I model of substitution, as determined by the Smart Model Selection in PhyML. The sequence DQ001069.1 (Genotype D) was used as reference sequence. The sequences from the genotypes D, L/D, and L, are highlighted in blue, red, and black, respectively. A thousand replicates were used to test the support given by the data to the clusters of the tree topology, and bootstrap values >70 were considered significant.

Figure 5: Schematic representation of the recombination event in sequences classified as L/D hybrid. The dataset containing the forty sequences were submitted to the RDP4 software program. KY985361.1 from Haiti, 2014, was identified as an L/D hybrid between the major parent KY618133.1 of Genotype L, Brazil, 1988 (red) and the minor parent KY618131.1 of Genotype D, Brazil, 1978 (blue). The other hybrids, KX496990.1

(Haiti, 2015) and KT818520.1 (Brazil, 2014) have their L major parents represented in orange and the D minor parents represented in blue. The image was generated employing GIMP 2.10.20 (www.gimp.org).

FIGURES

Figure 1

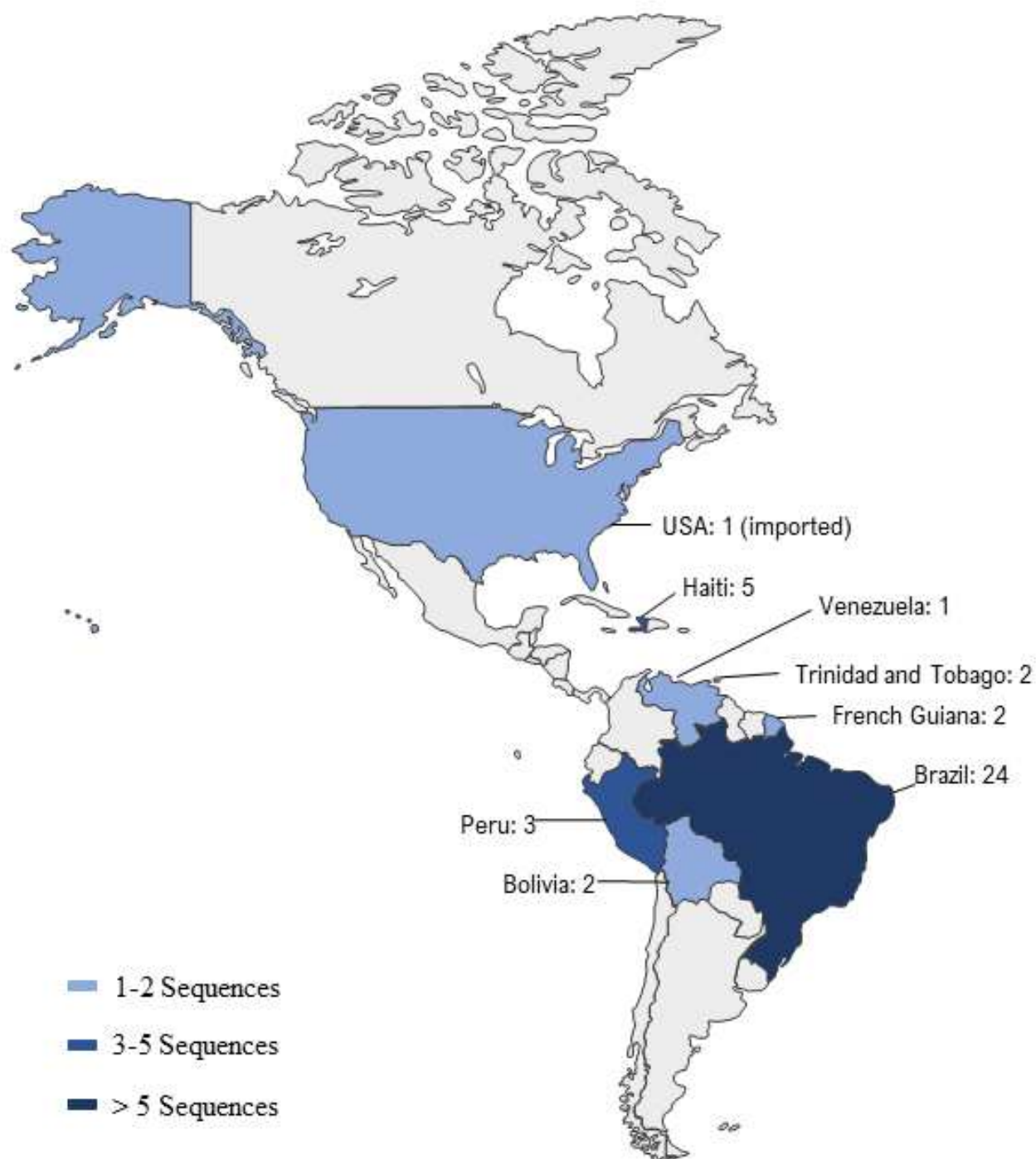


Figure 2

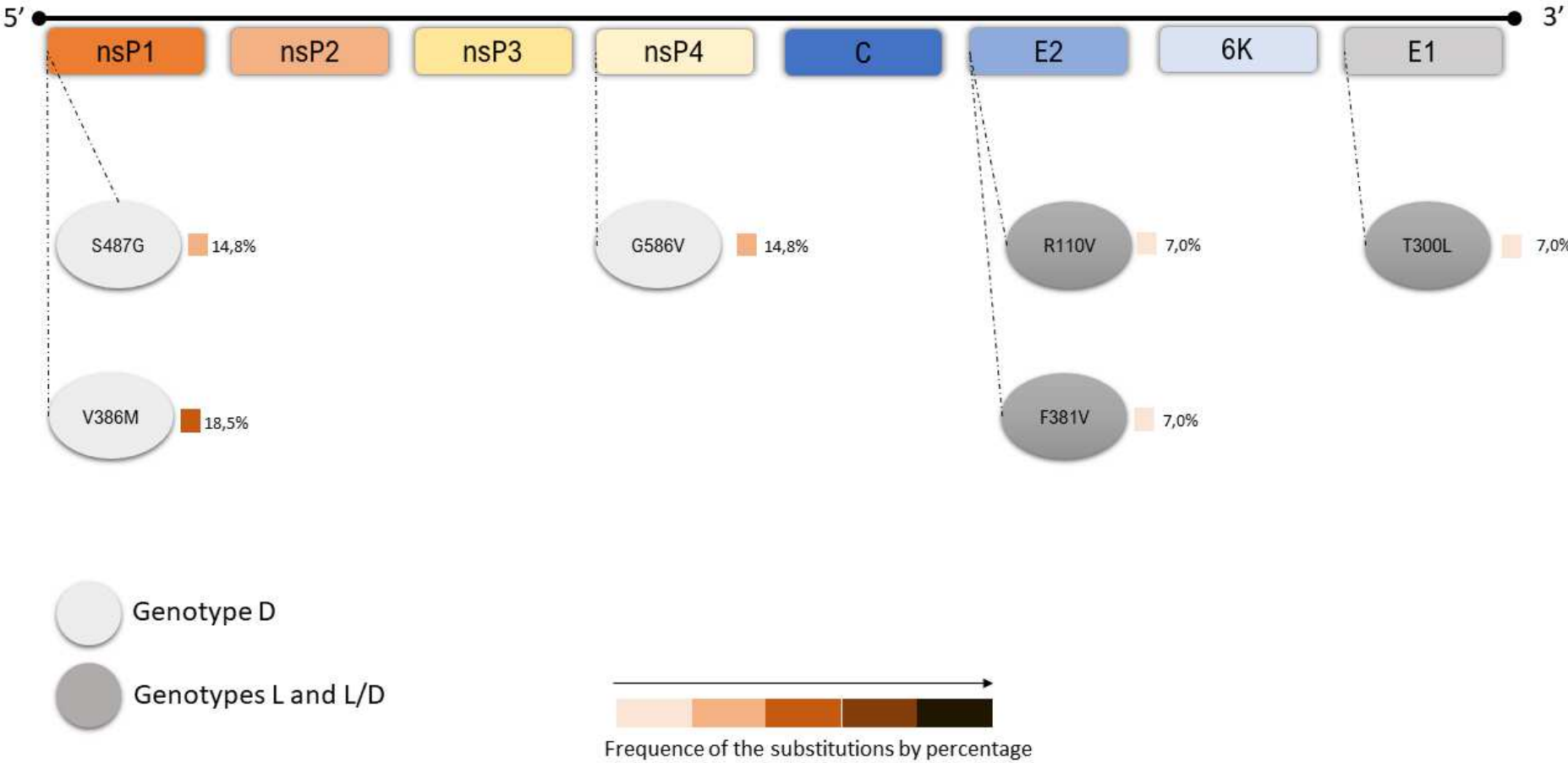


Figure 3

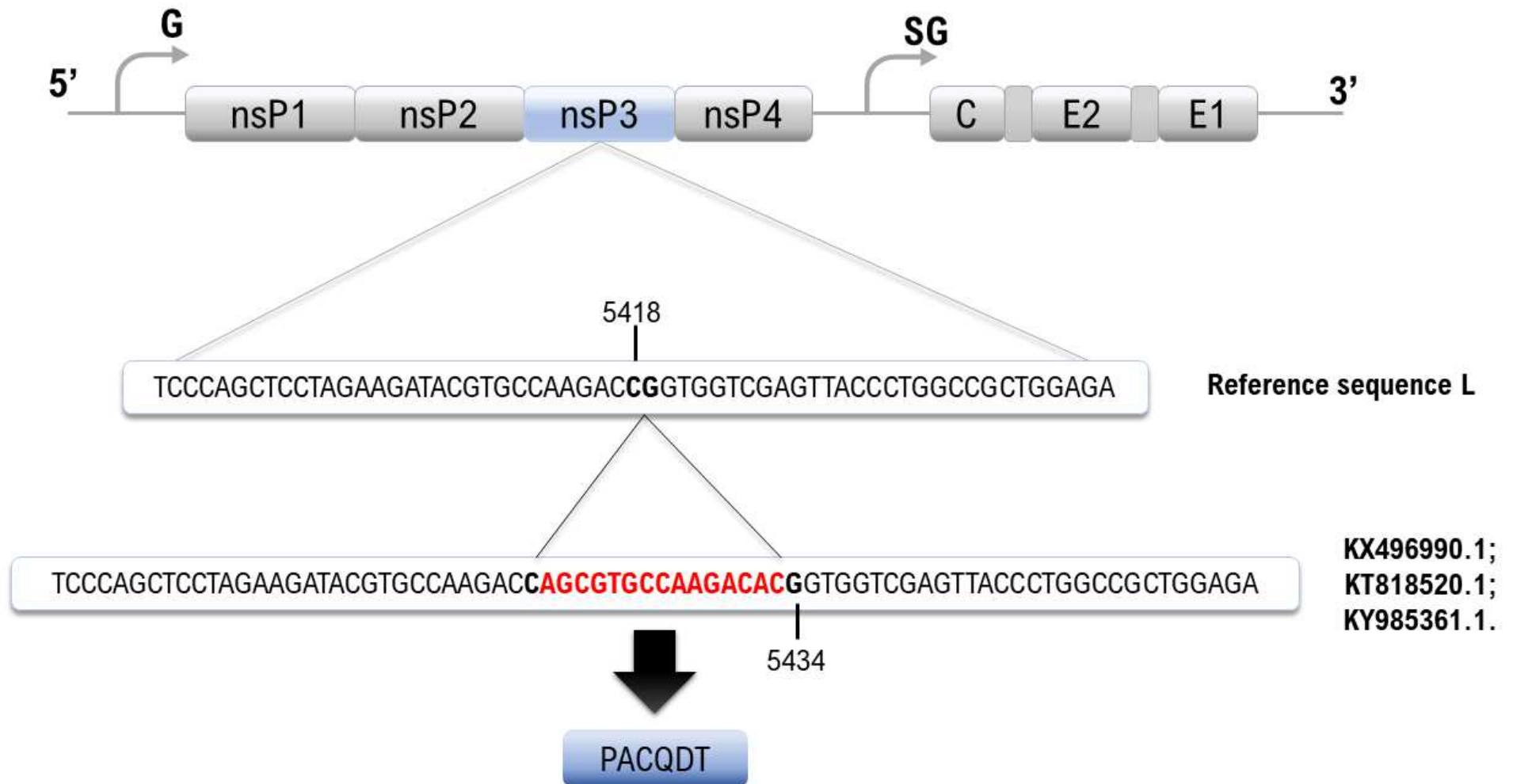


Figure 4

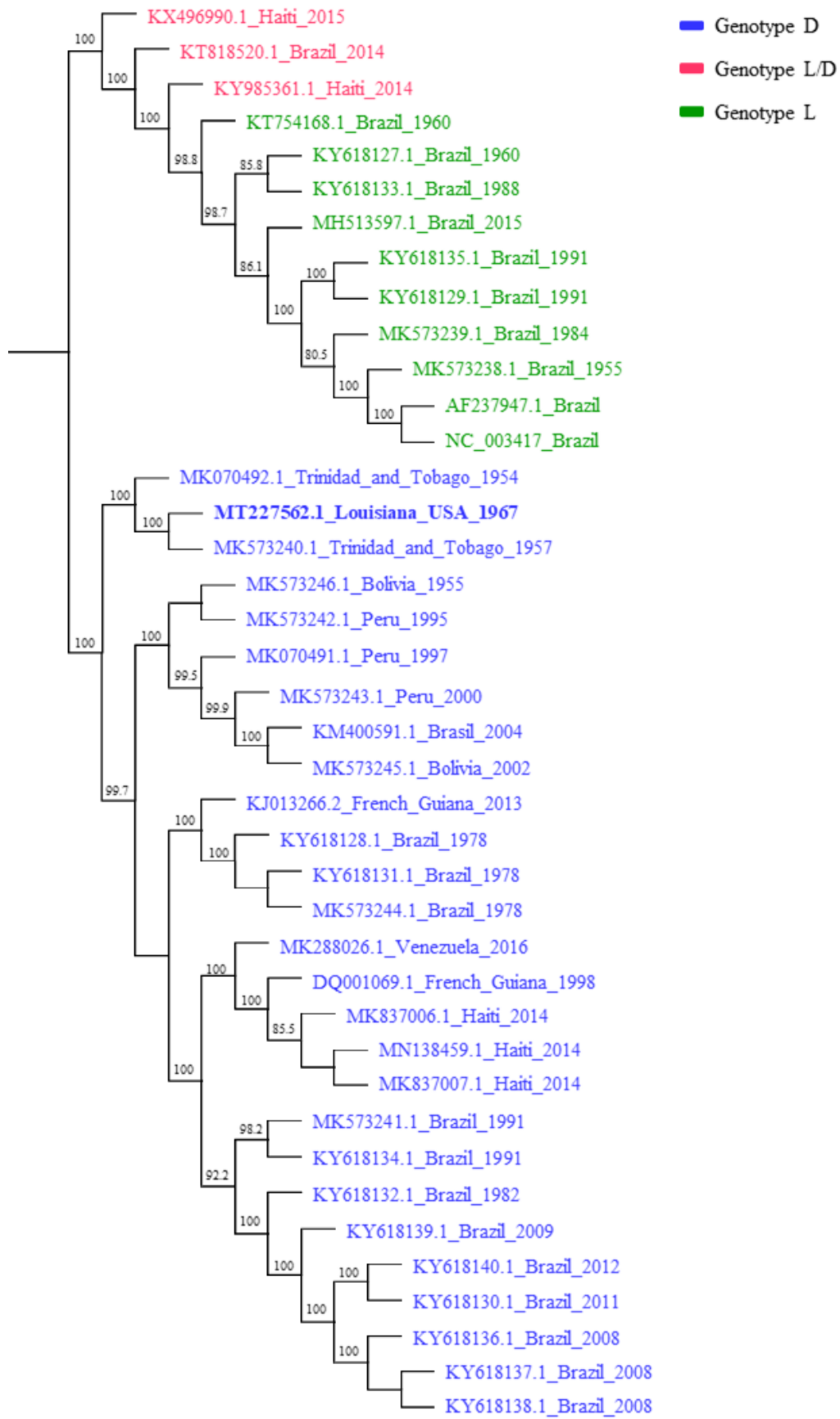


Figure 5

