Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

John Shen Date

Molecular Epidemiology of a Type 4 Dengue Virus Outbreak in Paraguay, 2019-2020

By

John Shen

MSPH

Global Epidemiology

_____

Benjamin Lopman, MSc, PhD

Committee Chair

_____

Kristin Bratton Nelson, MPH, PhD

Committee Member

_____

Anne Piantadosi, MD, PhD

Committee Member

_____

Jesse Waggoner, MD

Committee Member

Molecular Epidemiology of a Type 4 Dengue Virus Outbreak in Paraguay, 2019-2020

By

John Shen

B.S.

Boston College

2018

Thesis Committee Chair: Benjamin Lopman, MSc, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Global Epidemiology
2022

Abstract

Molecular Epidemiology of a Type 4 Dengue Virus Outbreak in Paraguay, 2019-2020

By John Shen

Background: Dengue virus type 4 (DENV-4), first detected in Paraguay in 2012, had persisted in the country as a non-predominant type. In 2019-2020, DENV-4 became predominant and caused an outbreak in Paraguay that spanned two waves: a smaller first wave in mid-2019 and a larger second wave between late 2019 and early 2020. We investigated the molecular epidemiology of the DENV-4 outbreak to understand its origin and evolution.

Methods: Patients with suspected dengue were enrolled in a cross-sectional study between January 2019 and March 2020 from two clinical facilities in Asunción, Paraguay. We selected DENV-4 positive samples for complete viral genome sequencing and phylogenetic analysis. Epidemiologic metadata were collected through a survey. Logistic regression models were run to find potential single nucleotide polymorphism (SNP) associations with disease severity.

Results: We sequenced complete DENV-4 genomes from 61 patients, largely concentrated in the Greater Asunción metropolitan area. Phylogenetic analysis revealed that outbreak viruses from 2019-2020 belonged to genotype II. The samples sequenced in this study were closely related to a small number of DENV-4 viruses detected in Paraguay in 2018, differing by only two synonymous mutations in NS3 and NS5. The shared ancestor between our Paraguay sequences and the closest sequence outside of Paraguay (Brazil 2013) dated to October 2010 (95% HPD: Mar 2010-May 2011). Among the outbreak sequences, we observed two clades, which diverged from a common ancestor in August 2017 (95% HPD: May 2017-Nov 2017), differed by three nucleotides (synonymous mutations in the E, NS3, and NS5 proteins), and were not separated by time or geography. We did not find a significant difference in the percentage of severe dengue cases between the clades ($p=0.70$), nor any SNPs associated with dengue severity.

Conclusions: Overall, our results suggest that a lineage of DENV-4 may have been circulating undetected in South America since 2010. We did not find evidence for substantial viral genomic differences between outbreak and pre-outbreak DENV-4 sequences, or between the first and second wave. A larger study with unbiased DENV sampling may determine if observed SNPs affect clinical manifestation of dengue.

Molecular Epidemiology of a Type 4 Dengue Virus Outbreak in Paraguay, 2019-2020

By

John Shen

B.S.

Boston College

2018

Thesis Committee Chair: Benjamin Lopman, MSc, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Global Epidemiology
2022

Acknowledgements

This thesis is made possible only with the continued guidance and support from all members of my committee. I would like to thank Dr. Benjamin Lopman for his insight on the epidemiologic methods in this project. I would also like to acknowledge Dr. Kristin Bratton Nelson for her contribution to the writing. Drs. Anne Piantadosi and Jesse Waggoner were pivotal to my development in phylogenetic analysis and molecular epidemiology methods, and I appreciate their supervision and mentorship throughout this project. Many thanks to Autum Key for her support in genetic sequencing and Alejandra Rojas, Fátima Cardozo, Cynthia Bernal, and César Cantero from the Health Sciences Research Institute at the National University of Asunción, San Lorenzo, Paraguay for case data acquisition and viral sample collection.

Last but certainly not least, I would like to thank all my friends and colleagues at Rollins for their love, energy, and constant support. I owe it to them for getting through the past two wild, unpredictable years.

Table of Contents

Introduction

Dengue is a mosquito-borne infectious disease affecting many tropical and subtropical regions of the world, which can spread rapidly in a community and for which there is no specific treatment (Gúzman & Kouri, 2002). The causative agent, dengue virus (DENV), is from a genus of single-stranded positive-sense RNA viruses known as *Flaviviruses* and consists of four main types (DENV-1 to 4) which have all originated from sylvatic Southeast Asian strains (Wang et al., 2000). Outcomes of DENV infection range from asymptomatic or subclinical infection to dengue fever to severe dengue, a potentially life-threatening hemorrhagic and capillary leak syndrome (Diamond & Pierson, 2015). Prior infection by a heterologous type confers some protection against a secondary infection by DENV-4. However, heterologous protection wanes over months to a few years, and DENV re-infection is the most important risk factor for severe dengue (Halstead, 2007). Dengue was first reported in Paraguay in 1989-1990 with epidemics of types 1, 2, and 3 reappearing every few years thereafter (Rojas et al., 2021). Dengue type 4 (DENV-4) first appeared in 2012, but persisted as a non-predominant type (Ramos-Castañeda et al., 2017). The most recent dengue virus outbreak in Paraguay started during the DENV season of 2018-2019 and extended into 2020. What is unique about this outbreak is that it was the first outbreak caused primarily by DENV-4, which was rarely seen in the country previously and is the least common type globally. The sudden outbreak of DENV-4 in Paraguay demonstrated the significant health impact that the virus can cause in a susceptible population. Indeed, this outbreak resulted in 60,891 confirmed cases (including 73 deaths) by early 2020, far exceeding epidemiologic trends in Paraguay for dengue outbreaks in prior years according to the July 2020 situation report published by the International Federation of Red Cross (IFRC, 2020). The

epidemic came in two waves, the first peaking in June 2019 and the second, larger wave peaking in December 2019-January 2020, associated with the capture of more severe cases.

The primary aim of this project is to characterize the source, timing, and genetic diversity of DENV-4 in the outbreak. The secondary aim is to determine if there are any single nucleotide polymorphisms (SNPs) that might be associated with more severe disease. Addressing the first aim, we use phylogenetic analysis to assess the origin of the DENV-4 outbreak, evaluate how viruses in the recent outbreak relate to earlier lineages in South America, and determine whether a clade shift or specific genetic mutations were associated with the substantial increase in cases (compared to the first wave) in Asunción, Paraguay's capital city, and the Central Department in late 2019 to early 2020. An understanding of this information can support surveillance and preparation for future outbreaks.

The second aim is to assess the association between viral SNPs and disease severity. Previous studies have suggested the potential involvement of mutations in three DENV nonstructural protein regions, NS1, NS3, and NS5, in DENV pathogenesis (Masrinoul et al., 2011; Delgado-Enciso et al., 2018). The NS1 and NS3 proteins are known targets for humoral immune responses, can thwart antiviral immune responses, and may be associated with severe clinical manifestations (Plaszczyca et al., 2019). NS5 is involved in viral replication. In addition, mutations in the envelope (E) protein may enhance host cell receptor binding, fusion, and entry (Fahimi et al., 2017). Our analysis takes into account other predictors of severe dengue, including age, pre-existing diabetes, prior infection with dengue, and renal disease (Tsheten et al., 2021) to understand the SNP-specific effect on dengue severity.

Materials and Methods

Ethical considerations

This study was reviewed and approved by the IICS Scientific and Ethics Committee (P38/2020) and the Emory Institutional Review Board (study 00110736). Written informed consent was obtained from all subjects. Children older than six years of age provided assent.

Study samples

Samples were collected between January 2019 and March 2020 at two sites in Asunción, Paraguay: Hospital Central of the Instituto de Previsón Social (IPS), an inpatient facility, and the Instituto de Investigaciones en Ciencias de la Salud, an outpatient clinical location. The individuals included resided in five of the 17 departments in Paraguay: Alto Paraguay, Alto Paraná, Boquerón, Central, and Paraguarí. Participants in the parent study were clinically diagnosed with an acute suspected arboviral illness compatible with dengue, and samples were selected for sequencing from individuals with confirmed infections with dengue virus serotype 4 (DENV-4) by reverse transcriptase-polymerase chain reaction (RT-PCR) and a Ct value < 35. Samples were further selected to represent the distribution of all cases based on month of collection, city of residence for the participant, and severity of the clinical illness according to the 2009 World Health Organization guidelines (dengue without warning signs, dengue with warning signs, and severe dengue; WHO, 2009). All samples from severe dengue cases were selected to maximize data from these participants. Epidemiologic metadata—demographic, diagnostic, and medical history information about each patient—was collected along with the samples.

Sequencing

A subset of 61 Paraguay DENV-4 RNA specimens underwent DENV full genome sequencing as follows: Extracted total nucleic acid underwent heat-labile dsDNase treatment (ArcticZymes, Tromso, Norway). cDNA was made using random hexamer primers (Fisher/Invitrogen) and SUPERSCRIPT III RT (Fisher/Invitrogen) for first strand synthesis, and New England Biolabs reagents for second strand synthesis, without amplification. Sequencing libraries were fragmented and indexed using the Nextera XT DNA Library Prep kit (Illumina) with dual indexes and 16 cycles of PCR. Libraries were quantified using the KAPA universal complete kit (Roche), pooled to equimolar concentration, and sequenced on a MiSeq with paired-end 150-bp reads (Illumina). As a negative control, water was included with each batch of samples starting from DNase. As a positive control, *in vitro* transcribed ERCC spike-ins (NIST) were added to each sample prior to cDNA synthesis.

Selection of reference sequences

Complete DENV-4 genomes were downloaded from the Virus Pathogen Database and Analysis Resource (ViPR, www.ViPRbrc.org) as reference sequences for phylogenetic analysis. These were MAFFT aligned with our Paraguay DENV-4 sequences using Geneious Prime (Biomatters, Inc., San Diego, CA, USA). Untranslated regions in the 5' and 3' ends were trimmed, resulting in 690 sequences of length 10,164 base pairs (bp). The following steps were taken to refine the dataset and downsample with representation across geography and time. Seventeen sequences, excepting sequences from Paraguay, were removed due to low coverage (>10% bp were ambiguous or missing). Four sequences were removed because they were collected after our Paraguay samples. Twenty-eight specimens with identical sequences, year of collection, and country of origin to another specimen were removed. Since the Paraguay DENV were most closely related to the 4-II genotype, 365 background sequences were filtered out for being

associated with one of the other three genotypes (4-I, 4-III, 4-IV sylvatic) described by the ViPR computational method and genotyping/subtyping tool (ViPR, 2022). Fourteen sequences were removed because either year or country of collection was missing. Finally, our downsampling scheme restricted sequences to one per year and country for contextual samples outside of the Americas, ten per year and country within the Americas but excluding Paraguay, and all Paraguay samples. At the beginning of an exploratory maximum likelihood (ML) tree run, a composition chi-square test was performed by IQ-TREE for every sequence to test for homogeneity of character composition (Trifinopoulos et al., 2016). One 2018 sequence from Paraguay, MT040677, failed this composition test because it significantly deviated from the mean composition in the alignment and depicted an unusually long branch topology in the resulting tree. Thus, this sequence was considered misclassified and removed.

Phylogenetic analysis

Phylogenetic trees were constructed with the IQ-TREE web server (an online interface for IQ-Tree software package version 1.6.12) (Trifinopoulos et al., 2016) using the ML method, with ultrafast bootstrap (UFBoot) approximation to compute support for the phylogenetic structure. The ModelFinder tool (Kalyaanamoorthy et al., 2017) was run to select the nucleotide substitution model most appropriate for our data, considering the Bayesian information criterion for 88 DNA substitution models.

We explored the temporal structure of the molecular phylogeny using TempEst v1.5.1 (Rambaut et al., 2016) and 12 reference sequences with >0.01 distance from the best-fitting linear regression were excluded as outliers for possible low sequencing quality or misclassified dates. Given the variable specificity of dates, the midpoint of the Gregorian calendar, July 2nd,

was imputed if only the year was provided for a sequence, and the 15th (midpoint of a month) was imputed for sequences providing both year and month.

We inferred the geographic origin, time to most recent common ancestor (TMRCA), and spatial dynamics using a Bayesian Markov chain Monte Carlo (MCMC) statistical framework implemented through BEAUti/BEAST v1.10.4 (Bouckaert et al., 2014). We chose the following parameters on BEAST to compute the maximum clade credibility (MCC) trees: constant convalescence, generalized time-reversible (GTR) substitution model with 4-category gamma-distributed base substitution rate, two partitions into codon positions, and 100 million chain length. Two versions of the tree were compared on TRACER v1.7.2, one assuming a strict molecular clock and the other assuming an uncorrelated relaxed clock. The better molecular clock model was chosen using the tree with a joint mean whose absolute value was closer to zero. TreeAnnotator v1.10.4 (Bouckaert et al., 2014) was used to summarize the MCC trees, which were then visualized through the interactive Tree of Life v6 (iTOL, https://itol.embl.de/), an online tool for the display and annotation of phylogenetic trees. TMRCA was also visualized on FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Analysis of SNP effect on dengue severity

For exploratory data analysis, we calculated frequencies of clinical risk factors among cases with severe dengue and those without (cases with asymptomatic dengue with or without warning signs). To test if such differences in frequencies were statistically significant, Fisher's exact tests were used to compare the proportions between the two levels of dengue severity. The Welch two sample t-test compared average ages. Risk factors that differed between outcome groups were included into a logistic regression model with dengue severity as the main outcome and an SNP as the main exposure, such that we could adjust for their effects while determining if certain SNP

mutations increased or decreased risk of severe dengue. The community where an individual was infected may affect their DENV-4 variant as well as their age, host genetics and prevalence of diabetes, renal disease, and other clinical risk factors, so these risk factors represent potential confounders. Because location is not well defined (an individual could be infected either in their home community or where they work, for example), we can only adjust for their age and clinical risk factors to reduce any spurious associations between SNP and dengue severity that are unrelated to the phenotypic expression of nucleotide mutations. We focused on SNPs that result in non-synonymous mutations, which alter the viral protein in the key regions described earlier and are thus likely to have an impact on DENV pathogenicity. All analyses were performed using R Statistical Software (v4.1.2; R Core Team 2021).

Results

Descriptive analysis of severe and non-severe dengue groups

Overall characteristics of our Paraguay outbreak sample population, as well as characteristics stratified by dengue case classification, are shown in **Table 1**. Risk factors for severe dengue such as age, pre-existing diabetes, chronic kidney disease, and prior dengue infection (Tsheten et al., 2021) were compared between the two case groups. The frequency of diabetes mellitus ($p=0.014$) was higher in the severe dengue group than the non-severe group. Average age is also greater for the severe dengue group than the non-severe group ($p=0.002$). Patients did not respond or did not know about previous dengue infection ($n=8$), leading to missing information in the severe dengue group.

Case distribution by symptom onset

Figure 1 describes the time distribution of our sequenced DENV cases by week of symptom onset. The first epidemic wave (n=16) was between February and August 2019, and the second epidemic wave (n=45) was between November 2019 and February 2020. There was almost a three-fold increase in number of cases in the second wave compared to the first.

Case distribution by geography

In both waves of the 2019-2020 outbreak, the cases were concentrated around the Greater Asunción metropolitan area, Central Department, Paraguay (**Figures 2A & 2B**). In the first wave, there were two cases outside of the Central Department (2/16; 13%): one in Ciudad del Este, Alto Paraná Department on the eastern border and the other in Puerto Casado, Alto Paraguay Department to the North. In the second wave, there were five cases outside the Central Department (5/45; 11%): three in neighboring Paraguarí Department and two in Boquerón in the Northwest.

Maximum likelihood tree

Following the filtering steps outlined in the **Methods**, 139 reference and 61 newly-generated DENV-4 sequences from Paraguay were in the final dataset. The reference sequences were from samples collected between 1956-2018 and included countries in Asia and the Americas. For our nucleotide substitution model, we selected the GTR model with a 4-category gamma-distributed site rate. The resulting ML tree was rooted on the oldest strain, KR011349 (GenBank), from Philippines in 1956. **Figure 3** shows a magnified section of the phylogenetic tree that contains our outbreak samples (highlighted in red and purple). The closest reference sequence, HVE159 (orange), was collected by our team in 2018 in Paraguay and sequenced for this study. Other 2018 Paraguay references from the ViPR database (unhighlighted), were more closely related to

our outbreak viruses than any other reference. The closest non-Paraguay reference was KP188564 from Brazil in 2013 (gray) at a genetic distance of 0.005 substitutions per site from the imputed common ancestor of all Paraguay DENV-4-genotype II. The ML tree also divides our sequences into two major clades, henceforth called clade A (n=24; red) and clade B (n=37; purple). This grouping is supported by 100% of the bootstrapped trees. Three SNPs separate the two clades, one SNP in the E protein and two in nonstructural proteins (NS3 and NS5). All three SNPs resulted in synonymous mutations. Some lower-level branches are colorized in red or orange, a result of not having enough informative genetic differences between such sequences to determine their evolutionary relationship.

Temporal structure of DENV-4 evolution

A linear regression using the residual mean squared function was fit on TempEst (r = 0.97) to assess if there is sufficient genetic change between sampling times to reconstruct a statistical relationship between genetic divergence and time. A high correlation coefficient with small residuals indicates that evolution could be represented by a Bayesian strict molecular clock model, which assumes that all branches of the tree have the same rate of evolution. The validity of this assumption was tested against an uncorrelated relaxed clock model that assumes the substitution rates of individual branches are drawn from a lognormal distribution. We determined the strict clock model to be more appropriate for this study. We also estimated the average mutation rate of DENV-4 in our phylogeny to be 7.9×10^{-4} mutations per site per year. This approximates the average rate for DENV for all serotypes at 7.5×10^{-4} mutations per site per year (Dang et al., 2020), which also includes the slower sylvatic cycle.

Maximum clade credibility tree

The resulting MCC tree is shown in **Figure 4**. The outbreak sequences are divided into the same clades as before. TMRCA analysis revealed that our outbreak samples diverged from the 2018 Paraguay sequences (excluding HVE159) in July 2017 (95% HPD: Mar-Oct 2017) and differed by only two nucleotides: synonymous mutations in NS3 and NS5. Although the HVE159 sequence collected by our group in 2018 switched clade membership, from clade A in the ML tree to clade B in the MCC tree, the posterior probability in the MCC tree was only 0.35, suggesting weak support. It is thus unclear whether all outbreak sequences were derived from a common ancestor with HVE159, or only the clade B outbreak sequences. Again, the KP188564 Brazil sequence from 2013 appears to be the closest non-Paraguay sequence to our outbreak samples. TMRCA analysis dated the divergence between the Paraguay DENV-4 sequences and KP188564 to October 2010 (95% HPD: Mar 2010-May 2011), suggesting prolonged undetected circulation of this DENV-4 lineage, most likely within South America. To investigate if there were any intermediary sequences collected between 2010 and 2018 that might provide better resolution, the reference sequence search was expanded to include all partial DENV-4 genomes with fully sequenced ~1400 bp E proteins between those years. After aligning these E gene reference sequences to our outbreak sequences and constructing a ML tree, no sequences more closely related to the Paraguay sequences than KP188564 were found.

According to the MCC tree, severe dengue cases did not appear to cluster together. Nor were they differentially distributed by clade: four cases (4/24; 17%) are found in clade A and four in clade B (4/37; 11%) (p=0.70). Similarly, no genetic clustering was observed for cases that had illness onset in the first epidemic wave, between February and August 2019, versus the second wave, between November 2019 and March 2020. The waves were distributed among both clades with no clear pattern.

SNP effect on dengue severity

We identified four SNPs where both variants appear in the severe and non-severe dengue groups (there were no SNPs that were exclusively one group or the other) that result in nonsynonymous mutations in the E, NS1, 2K, and NS5 genes: U1534A, U3058C, A6697G, and A8964C respectively. The minor variant frequencies are low. In the E and NS1 proteins, the minor variant appeared one time (1.9%) in the non-severe group and once in the severe dengue group (12%). In the 2K and NS5, the minor variant appeared four times (7.5%) in the non-severe cases and once (12%) among the severe cases.

A logistic regression was performed to determine the odds of dengue among those with compared to without a mutation (the minor variant of an SNP). Age and pre-existing diabetes mellitus were included in the model as covariates because they differed between the two outcome groups, as noted earlier in this section. The results are shown in **Table 2** below.

While cases with the minor variant of mutations on the E and NS1 proteins have a higher odds of severe dengue than those without these mutations, adjusting for covariates in the model, the difference is not statistically significant. In the same way, cases with the minor variant on the 2K and NS5 have lower odds of severe dengue than individuals without, but these differences are also not significant. Therefore, we cannot reject the null hypothesis that the four mutations are not associated with severe clinical manifestations of DENV infection.

Discussion

Phylogenetic analysis revealed that outbreak viruses from 2019-2020 belonged to genotype II, which had been responsible for outbreaks in South America and the Caribbean. It also showed that the outbreak sequences were very closely related to the small number of

DENV-4 sequences already circulating in Paraguay in 2018. Thus, importation of a new DENV-4 lineage did not explain the large outbreak, which resulted in a switch in predominant type from DENV-1 in the 2018 outbreak, and virus evolution is unlikely to explain why DENV-4 suddenly became predominant in the country after circulating for seven years. A paper on DENV-4 in Puerto Rico suggests an alternative explanation for the upsurge. The authors noticed that viral diversity increased before the appearance of outbreaks, and this acceleration in DENV evolution may be attributed to positive selection of NS2A, a gene contributing to replication antigenicity and efficiency (Bennett et al., 2003; Bennett et al., 2010). While no mutation in this particular gene (or any other nonsynonymous mutations) was observed for the 2018 Paraguay lineages, this could be a result of the lack of sequenced genotype II DENV-4 viruses in Paraguay before 2018. In fact, there is scarce information on DENV-4 in South America after 2013 in Brazil. This limits our ability to detect key mutations which may have increased the viral diversity leading up to the 2019-2020 outbreak.

Furthermore, we did not observe genetic differences between the two waves in 2019-2020, so it is unlikely any specific mutation event led to a more transmissible DENV-4 lineage in late 2019 to produce far more cases than in the first 2019 wave. One possibility is that viral replication increased in the *Aedes* mosquito vector with an increase in ambient temperature in December and February (Morin et al., 2013). All eight severe dengue cases were detected in the second wave; however, this may be attributed to the greater number of specimens in the second wave sequenced overall, allowing for detection of more severe dengue by chance, rather than the emergence of a more virulent strain. This is supported by the fact that the percentages of severe dengue cases did not significantly differ between the two major clades. Clades A and B also did not vary by location, indicating that these two lineages were not geographically isolated.

Absence of a clear connection between clade genetics and greater incidence of DENV4 in the second wave suggests that if there was a mutation that stimulated DENV4 circulation and diversity, it may have developed before the common ancestor of clades A and B, leading to the proliferation of both clades over time.

In addition to the E, NS3, and NS5 proteins, a nonsynonymous mutation was also found in the 2K region on DENV-4. Previous literature does support the interaction between NS1 and the NS4A-2K-4B cleavage as a potential contributor to the RNA replication and amplification process (Plaszczyca et al., 2019), so this mutation was added to our analysis. The absence of significant associations between SNPs and dengue severity was anticipated for several reasons. First, the study was conducted with a small sample size, with only eight patients who were classified as cases of severe dengue. With a sample this small, one can expect high variance in results based on the people who were selected into the sample, reflected by the large confidence intervals. This limitation affects our ability to infer whether the differences between severe and non-severe dengue cases in **Table 1** reflect true differences among cases of DENV4 in the Paraguay outbreak. Consequently, the model chosen to investigate SNP associations may be misspecified if some of the dropped covariates would have remained in the model in a larger study. A small sample size also means the study is not powered to make inferences on true SNP associations, should they exist.

Secondly, because sequenced viruses came from a convenience sample from two health facilities, we are limited in our ability to generalize our findings to all symptomatic cases, severe or otherwise. By enrolling patients who were admitted to these facilities, there is a systematic bias to select only those patients with worse outcomes, whereas those with milder symptoms would be less likely to seek medical care. Geographically, we are also limiting our target

population to people living near the two facilities in Asunción, which explains the spatial clustering around Asunción. A future study seeking to find SNPs associated with severe symptomatic dengue as opposed to non-severe symptomatic dengue in Paraguay would require random genetic sampling of the entire population, across all departments regardless of reported symptoms. This approach would also be able to capture the estimated 50-85% of asymptomatic cases (Duong et al., 2015), allowing for the identification of mutations that may distinguish asymptomatic versus symptomatic cases. It remains an open question whether systematic differences in viral genetics between those causing asymptomatic compared to symptomatic dengue exist, as differences in case presentation may be more attributed to presence of host genes that confer protection against clinical dengue (Yeo et al., 2014).

Conclusions

Overall, the results of this study suggest that the lineage of DENV-4 that caused the 2019-2020 outbreak may have been circulating undetected in South America since 2010, when it diverged from the sample sequenced in Brazil in 2013. We did not find evidence of recent importation as the main driver for the 2019-2020 DENV-4 outbreak, and it is more likely that they evolved from strains already in Paraguay in 2018. There was also no evidence for substantial genomic differences between sequences in the first and second waves of this epidemic. Finally, further research on a larger sample is needed to confirm or reject a causal relationship between the four identified SNPs and severe dengue. This larger sample should come from random sero-sampling of healthy controls as well as symptomatic cases.

References

Bennett SN, Holmes EC, Chirivella M, et al. Selection-driven evolution of emergent dengue virus. Molecular Biology and Evolution. 2003 Oct;20(10):1650-1658. DOI: 10.1093/molbev/msg182. PMID: 12832629.

Bennett, S. N., Drummond, A. J., Kapan, D. D., Suchard, M. A., Muñoz-Jordán, J. L., Pybus, O. G., Holmes, E. C., & Gubler, D. J. (2010). Epidemic dynamics revealed in dengue evolution. *Molecular biology and evolution*, *27*(4), 811–818. https://doi.org/10.1093/molbev/msp285

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS computational biology. 2014;10(4). 10.1371/journal.pcbi.1003537

Cuellar, C. M., Lovera, D., Merlo, O., & Arbo, A. (2020). Impacto económico del dengue en Paraguay [Dengue economic impact in Paraguay]. *Revista chilena de infectologia: organo oficial de la Sociedad Chilena de Infectologia*, *37*(4), 356–361. https://doi.org/10.4067/S0716-10182020000400356

Dang, T.T., Pham, M.H., Bui, H.V. *et al.* Whole genome sequencing and genetic variations in several dengue virus type 1 strains from unusual dengue epidemic of 2017 in Vietnam. *Virol J* **17,** 7 (2020). https://doi.org/10.1186/s12985-020-1280-z

Delgado-Enciso, I., López-Lemus, U.A., Valcarcel-Gamiño, J.A., Rodriguez-Sanchez, I.P., Valle-Reyes, S., Martinez-Fierro, M.L. ... Espinoza-Gómez, F. (2018). Dengue virus-1 NS5 genetic variant associated with a severe clinical infection: Possible reduction of the innate immune response by inhibition of interferon type 1 and the Janus kinase-signal transducer and activator of transcription signaling pathway. International Journal of Molecular Medicine, 41, 2263-2269. https://doi.org/10.3892/ijmm.2018.3395

Diamond, M. S., & Pierson, T. C. (2015). Molecular Insight into Dengue Virus Pathogenesis and Its Implications for Disease Control. *Cell*, *162*(3), 488–492. https://doi.org/10.1016/j.cell.2015.07.005

Duong, V., Lambrechts, L., Paul, R. E., Ly, S., Lay, R. S., Long, K. C., Huy, R., Tarantola, A., Scott, T. W., Sakuntabhai, A., & Buchy, P. (2015). Asymptomatic humans transmit dengue virus to mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), 14688–14693. https://doi.org/10.1073/pnas.1508114112

Fahimi, H., Mohammadipour, M., Haddad Kashani, H., Parvini, F., & Sadeghizadeh, M. (2018). Dengue viruses and promising envelope protein domain III-based vaccines. *Applied microbiology and biotechnology*, *102*(7), 2977–2996. https://doi.org/10.1007/s00253-018-8822-y

Guzmán, M. G., & Kourí, G. (2002). Dengue: an update. *The Lancet. Infectious diseases*, *2*(1), 33–42. https://doi.org/10.1016/s1473-3099(01)00171-2

Halstead S. B. (2007). Dengue. *The Lancet*. 370(9599), 1644-1652. https://doi.org/10.1016/S0140-6736(07)61687-0

International Federation of Red Cross and Red Crescent Societies. (2020). *Paraguay: Dengue outbreak (MDRPY021) Final Report*. Reliefweb. https://reliefweb.int/report/paraguay/paraguay-dengue-outbreak-mdrpy021-final-report

Masrinoul, P., Diata, M. O., Pambudi, S., Limkittikul, K., Ikuta, K., & Kurosu, T. (2011). Highly conserved region 141–168 of the NS1 protein is a new common epitope region of dengue virus. *Japanese journal of infectious diseases*, *64*(2), 109–115.

Morin, C. W., Comrie, A. C., & Ernst, K. (2013). Climate and dengue transmission: evidence and implications. *Environmental health perspectives*, *121*(11-12), 1264–1272. https://doi.org/10.1289/ehp.1306556

Płaszczyca, A., Scaturro, P., Neufeldt, C. J., Cortese, M., Cerikan, B., Ferla, S., Brancale, A., Pichlmair, A., & Bartenschlager, R. (2019). A novel interaction between dengue virus nonstructural protein 1 and the NS4A-2K-4B precursor is required for viral RNA replication but not for formation of the membranous replication organelle. *PLoS pathogens*, *15*(5), e1007736. https://doi.org/10.1371/journal.ppat.1007736

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramos-Castañeda, J., Barreto Dos Santos, F., Martínez-Vega, R., Galvão de Araujo, J. M., Joint, G., & Sarti, E. (2017). Dengue in Latin America: Systematic Review of Molecular Epidemiological Trends. *PLoS neglected tropical diseases*, *11*(1), e0005224. https://doi.org/10.1371/journal.pntd.0005224

Rambaut A. FigTree 1.4.4. 2018. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016 Apr 9;2(1):vew007. doi: 10.1093/ve/vew007. PMID: 27774300; PMCID: PMC4989882.

Rojas, A., Moreira Soares, A., Mendoza, L. P., Acosta, M. E., Aria, L., Páez, M., Herebia, L., Vallejos, M. A., de Guillén, Y., & Aquino, V. H. (2021). Revisiting the dengue epidemic of 2011 in Paraguay: molecular epidemiology of dengue virus in the Asuncion metropolitan area. *BMC infectious diseases*, *21*(1), 769. https://doi.org/10.1186/s12879-021-06487-9

S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, and L.S. Jermiin (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 14:587–589. DOI: 10.1038/nmeth.4285

Sjoberg D, Whiting K, Curry M, Lavery J, Larmarange J (2021). "Reproducible Summary Tables with the gtsummary Package." *The R Journal*, **13**, 570-580. doi: 10.32614/RJ-2021-053, https://doi.org/10.32614/RJ-2021-053.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) *Nucl. Acids Res.* 44 (W1): W232-W235. doi: 10.1093/nar/gkw256

Tsheten, T., Clements, A.C.A., Gray, D.J. *et al.* Clinical predictors of severe dengue: a systematic review and meta-analysis. *Infect Dis Poverty* **10,** 123 (2021). https://doi.org/10.1186/s40249-021-00908-2

Virus pathogen database and analysis resource (vipr) - genome database with visualization and Analysis Tools. Virus Pathogen Database and Analysis Resource (ViPR) - Genome database with visualization and analysis tools. (n.d.). Retrieved March 7, 2022, from https://www.viprbrc.org/

Wang, E., Ni, H., Xu, R., Barrett, A. D., Watowich, S. J., Gubler, D. J., & Weaver, S. C. (2000). Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *Journal of virology*, *74*(7), 3227–3234. https://doi.org/10.1128/jvi.74.7.3227-3234.2000

World Health Organization. Dengue Hemorrhagic Fever: Diagnosis, Treatment, Prevention and Control. new ed. Geneva, Switzerland: World Health Organization; 2009.

Yeo, A. S., Azhar, N. A., Yeow, W., Talbot, C. C., Jr, Khan, M. A., Shankar, E. M., Rathakrishnan, A., Azizan, A., Wang, S. M., Lee, S. K., Fong, M. Y., Manikam, R., & Devi Sekaran, S. (2014). Lack of clinical manifestations in asymptomatic dengue infection is attributed to broad down-regulation and selective up-regulation of host defence response genes. *PloS one*, *9*(4), e92240. https://doi.org/10.1371/journal.pone.0092240

Figures and Tables

Table 1. Characteristics of DENV-4 infected case-patients in the study, overall and stratified by dengue classification: DSSA = dengue without warning signs, DCSA = dengue with warning signs, DG = severe dengue. Table created with R *gtsummary* package (Sjoberg et al., 2021)

Characteristic	**N**	**Overall**, N = 61[1]	**Not severe (DSSA or DCSA)**, N = 53[1]	**Severe dengue (DG)**, N = 8[1]
Gender	61			
Female		40 (66%)	36 (68%)	4 (50%)
Male		21 (34%)	17 (32%)	4 (50%)
Age (years)	61	39.2 (20.4)	35.9 (19.2)	60.9 (15.0)
Diabetes mellitus	61	5 (8.2%)	2 (3.8%)	3 (38%)
Chronic kidney disease	61	2 (3.3%)	1 (1.9%)	1 (12%)
Previous dengue infection	53	22 (42%)	22 (42%)	0 (NA%)

[1] n (%); Mean (SD)

Figure 1. Weekly symptom onset of sequenced DENV-4 cases. Note: alignment of bars, vertical gridlines, and axis labels on Monday weeks. Dates of symptom onset range from 27 February 2019 to 04 March 2020. Bars colored in red are sequences from Greater Asunción (Central Department), blue otherwise.

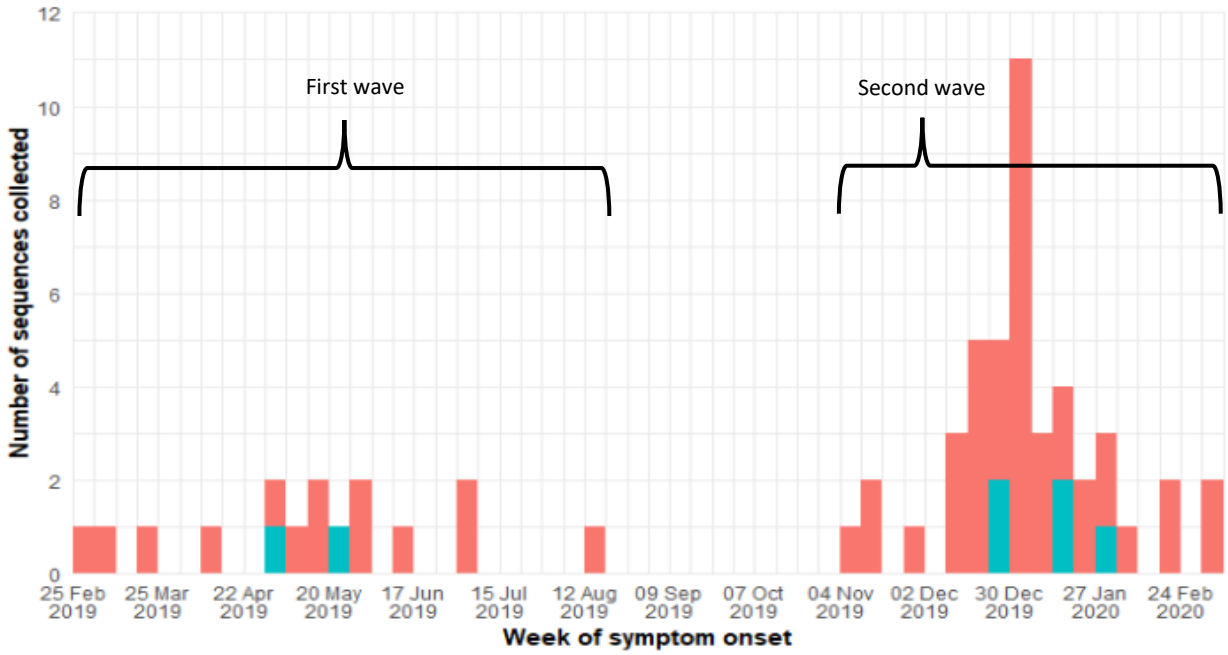

Figure 2. Geographic distribution of sequenced DENV-4 cases in Paraguay during the first (2A) and second (2B) epidemic waves. The red boxes (left) mark the boundaries of the magnified areas (right) of greater Asunción and Paraguarí. Visualized on Google Earth 9.159.0.0. (Google Inc., USA). Map based on a satellite image available in Google Earth (http://earth.google.com)

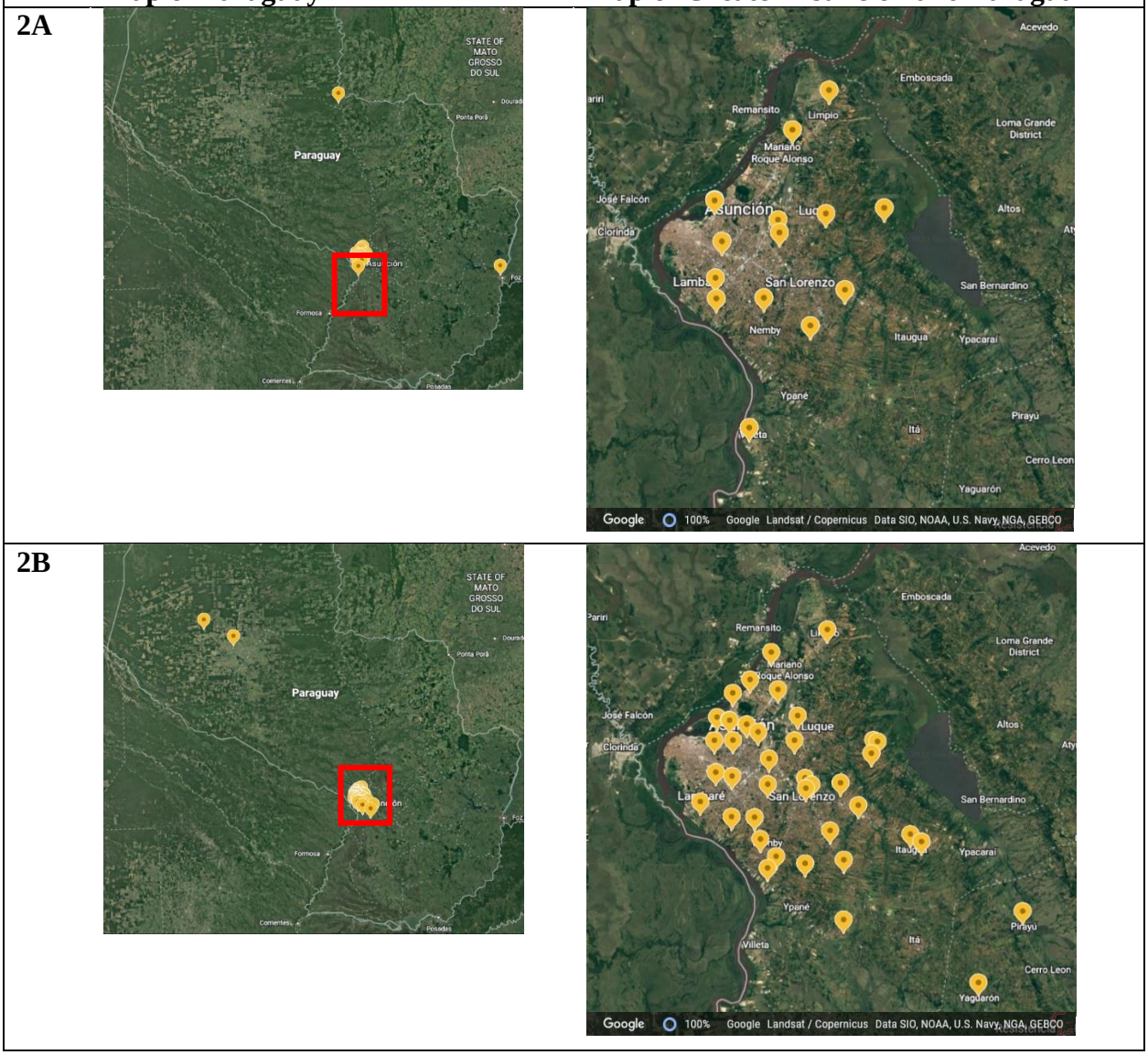

Figure 3. Maximum likelihood tree of DENV-4-genotype II sequences created with IQTree and visualized on iTOL. Magnified and cropped to show Paraguay 2019-2020 outbreak sequences and closest references. UFboot support displayed by coloring of branches (red = ~0% of bootstrap trees, little to no support; green = ~100%, extensive support). Highlighted labels: Brazil 2013 reference sequence (gray), closest Paraguay 2018 reference sequence (orange), Paraguay outbreak sequences in clade A (red), Paraguay outbreak sequences in clade B (purple).

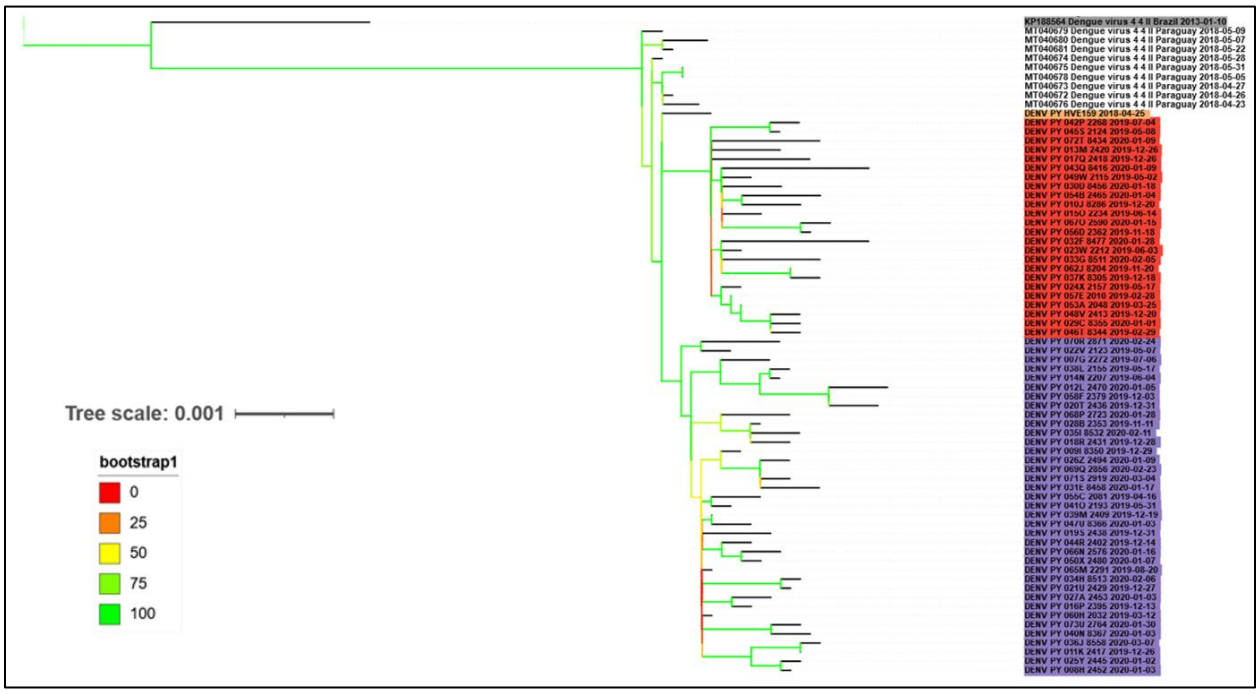

Figure 4. MCC tree of DENV-4-genotype II created with BEAST and visualized on iTOL. Magnified and cropped to show Paraguay 2019-2020 outbreak sequences and closest references. Cases classified as severe dengue is indicated with a red box to the right, else marked with an empty box. Years are indicated by the vertical lines and labelled on the bottom axis. Highlighted labels: Brazil 2013 reference sequence (gray), closest Paraguay 2018 reference sequence (orange), Paraguay outbreak sequences in clade A (red), Paraguay outbreak sequences in clade B (purple).

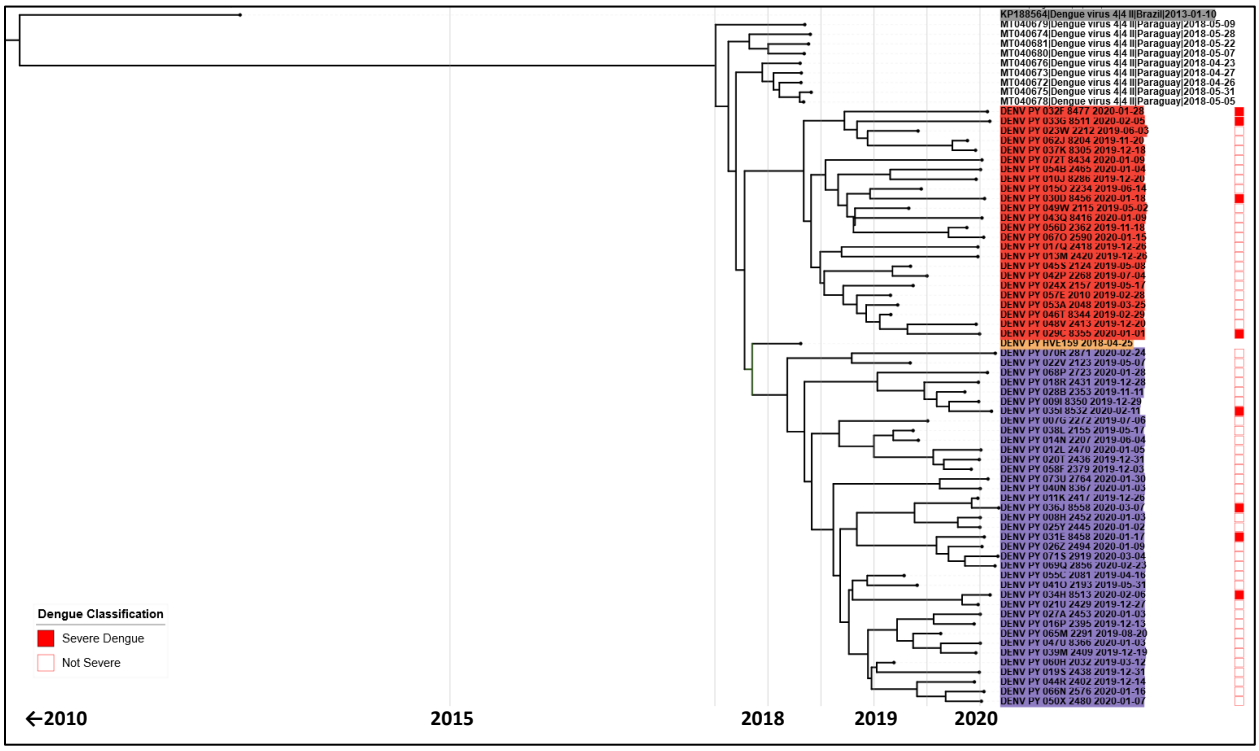

Table 2. Odds ratios of severe dengue among those with minor variant nucleotide compared to those without. OR = odds ratio, CI = confidence interval.

Mutation	***Location on DENV-4***	***Minor Variant Nucleotide***	***OR (95% CI)***
U1534A	E	A	16.1 (0.7, 386.6)
U3058C	NS1	C	16.1 (0.7, 386.6)
A6697G	2K	G	0.6 (0.0, 1214.3)
A8964C	NS5	C	0.9 (0.0, 20.3)