Title: Proposing a Systematic Lineage Classification Below the Genotype Level for Dengue Serotypes 1 and 2

James Siqueira Pereira[1], Gabriela Ribeiro[1], Vinicius Carius de Souza[1], Isabela Carvalho Brcko[1], Igor Santana Ribeiro[1], Iago Trezena Tavares De Lima[1], Svetoslav Nanev Slavov[1], Sandra Coccuzzo Sampaio Vessoni[1], Maria Carolina Elias[1], Alex Ranieri Jerônimo Lima[1,#]

[1] Center for Viral Surveillance and Serological Assessment (CeVIVAS), Butantan Institute, São Paulo, Brazil;

Corresponding author
Alex Ranieri J. Lima, Butantan Institute, Av. Vital Brasil, 1500, São Paulo, 05503-900, Brazil, alex.lima@fundacaobutantan.org.br

Abstract

Dengue virus (DENV), a mosquito-borne flavivirus, is causing a significant outbreak in Brazil. The recent surge in complete DENV genome sequences necessitates a standardized classification system for an improved understanding of viral dynamics and transmission patterns. Traditionally, DENV classification relies on serotypes and genotypes but lacks a consensus for sub-genotype classification. This hinders comprehensive analyses of viral diversity. We address this gap by proposing a novel lineage classification system for DENV using a semi-automatic workflow, leveraging the use of complete genome sequences to classify and re-evaluate DENV genetic diversity. This system offers a more granular classification scheme compared to current methods. The proposed hierarchical nomenclature, incorporating serotype, genotype, subgenotype, lineage, and sublineage, facilitates precise tracking of viral introductions and evolutionary events. This information might have crucial implications for public health interventions, enabling more targeted control strategies and improved monitoring of vaccine effectiveness during future outbreaks.

Keywords: Dengue virus (DENV), Lineage classification, Genomic surveillance, Public health

Introduction

Dengue fever, caused by the four Dengue virus (DENV) serotypes, stands as a paramount arboviral ailment globally. Manifestations range from mild febrile illness to severe hemorrhagic diathesis, also recognized as Dengue Shock Syndrome (Wang *et al.*, 2020). Transmitted by *Aedes* mosquito vectors, notably *Aedes aegypti* and *Aedes albopictus*, the disease exhibits an expanding geographical footprint, portending an inevitable surge in DENV-related morbidity and mortality rates worldwide. Notably, in 2023, the global toll encompassed over five million reported cases of dengue, accompanied by more than 5000 fatalities across 80 countries. The Americas Region bore the brunt, recording approximately 4.1 million cases, representing nearly 80% of the global burden (WHO, 2023). This high transmission rate has continued into 2024, with a total of 1,874,021 suspected cases of dengue reported spanning from epidemiological weeks 1 to 8. Out of these cases, 658,215 were lab-confirmed, 1,670 were classified as severe (0.1%), with 422 resulting in fatalities (case fatality rate 0.023%) (PAHO, 2024).

DENV belongs to the *Flaviviridae* family with a well-defined genomic structure: a single positive strand RNA of approximately 11 Kb. The genomic organization encompasses two untranslated regions (UTR) at the 5' and 3' ends, flanking an Open Read Frame (ORF), which serves as a template to encode a polyprotein composed of 3 structural proteins and 7 non-structural proteins. The structural proteins are the capsid (C), pre-membrane (PrM), and envelope (E), while the non-structural proteins are numbered as NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Nanaware *et al.*, 2021). There are four distinct DENV serotypes (DENV-1-4), along with a recently described fifth serotype (DENV-5) (Mustafa *et al.*, 2015). In addition to this classification, each serotype is divided into different genotypes, often defined as strains with less than 6% divergence at the junction between the E and NS1 coding regions (Rico-Hesse, 1990).

The elucidation of virus lineages and strains is pivotal for comprehending their evolutionary trajectories. Integrating this genetic information with diverse datasets, including geographical, epidemiological, and clinical parameters, furnishes invaluable insights into the disease dynamics (Grubaugh, 2022). Previous studies suggest that DENV lineages go extinct every 7-10 years, being replaced by new lineages, typically resulting in epidemics (Adams *et al.*, 2006; Nunes *et al.*, 2014; Lourenço *et al.*, 2018).

Nonetheless, the absence of a standardized lineage classification and nomenclature for DENV at subgenotype levels poses a significant challenge in longitudinally monitoring lineage substitutions. Moreover, communicating lineage-replacement events or identifying epidemic strains to stakeholders and public health agencies becomes problematic in the absence of an established lineage system. The utilization of the lineage classification system proposed by Rambaut et al., 2020 has successfully overcome those issues during the COVID-19 pandemic, as it helped to precisely track the virus variants across the globe for SARS-CoV-2 (Oude Munnink and Koopmans, 2023; WHO, 2024). This approach has since been adapted to propose lineage designation frameworks for other viruses, including Monkeypox, Rabies, and HRSV (Campbell *et al.*, 2022; Happi *et al.*, 2022; Goya *et al.*, 2024), leveraging the rationale behind its implementation.

In March 2022, the World Health Organization (WHO) launched the Global Arbovirus Initiative, a collaborative effort designed to unite key stakeholders in bolstering surveillance and preventative strategies against arboviruses (Balakrishnan, 2022). Leveraging existing genomic surveillance networks established for SARS-CoV-2, this initiative facilitated the real-time availability of comprehensive DENV genomes. With the escalation of the number of dengue cases came a pressing necessity to systematically organize and centralize this burgeoning genomic data, culminating in establishing the GISAID EpiArbo database in 2023 (Wallau *et al.*, 2023). Concurrently, novel bioinformatic tools emerged, geared towards impartially proposing and delineating new lineage classification systems for pathogens, adept at managing vast genomic datasets (Campbell *et al.*, 2022; Ha and Aylward, 2024; McBroome *et al.*, 2024)

Brazil has become a significant area of concern due to the rapid rise in dengue cases. During the first five epidemiological weeks of 2024, a substantial number of 455,525 dengue cases were reported in Brazil (PAHO and WHO, 2024). This concerning trend continued, with an additional 239,058 cases reported by epidemiological week eight (PAHO, 2024). Notably, while all four DENV serotypes have been identified in Brazil, in recent years there is also intensive circulation of DENV serotypes 1 and 2 (Figure S1). This regional dominance underscores the urgent need for tailored strategies aimed at classification and surveillance to manage and control dengue within the country's borders effectively. As such, understanding the dynamics of these prevalent serotypes beyond the genotype level within Brazil holds

significant implications for global dengue control efforts, necessitating a concerted focus on this geographic hotspot. Thus, we propose a lineage classification system for DENV serotypes 1 and 2 in this work, applying a semi-automatic workflow.

Methods

Retrieving and filtering the sequence dataset

The genomes of DENV-1 and DENV-2 were obtained from GISAID EpiArbo (https://www.gisaid.org/) selecting sequences with a minimum 10,000-nt, aiming to retrieve only genomes expected to be complete. Only DENV original passages and genomes obtained from the human host were selected. The DENV-1 and DENV-2 datasets contained sequences obtained up to January 2024 with initial collection dates between 1944 and 1976. To ensure comprehensive data coverage, we employed an in-house established script to assess the presence of ambiguous bases in each genome, setting a 10% cutoff. This strategy ensured that only genomes with at least 90% coverage would be retained. The metadata of the sequences used (including the GISAID accession IDs) are available in Supplementary files 1 and 2.

Phylogenetic tree reconstruction

Genomic sequences from the two filtered serotype datasets were aligned independently using the Augur align tool within the Augur program toolkit v24.1.0 (Huddleston *et al.*, 2021), employing the MAFFT v.7.520 algorithm (Katoh and Standley, 2013). Subsequently, maximum likelihood (ML) phylogenetic trees were constructed utilizing IQ-TREE v2.2.6 (Minh *et al.*, 2020). The trees were reconstructed employing the GTR+F+I+G4 nucleotide substitution model, selected by the ModelFinder application (Kalyaanamoorthy *et al.*, 2017) within IQ-TREE2. To assess the robustness of tree topology, 5000 UFBoot replicates were generated. Phylogenetic trees were dated using Treetime v0.11.2 (Sagulenko, Puller and Neher, 2018), integrated in Augur. Furthermore, multiple refinements were applied during this phase, including rerooting the tree using the midpoint method, inference of node confidence dates, stochastic resolution of polytomies, and pruning of genomes that exhibited deviation from the evolutionary rate (clock filter interquartile distance of 4). To automate the analysis workflow, a local Bash script was developed and is available on

the GitHub repository (https://github.com/alex-ranieri/denvLineages/blob/main/scripts/execPhyloDenv.sh).

Ancestral inference and identification of nucleotide mutations were conducted using Augur ancestral, while amino acid mutations were annotated utilizing Augur translate. For DENV-1, the GenBank sequences NC_001477.1 served as the reference, and for DENV-2, NC_001474.2 was utilized.

Lineage definition workflow

Initially, DENV-1 and DENV-2 lineages were defined using an automated heuristic approach implemented in Autolin (McBroome *et al.*, 2024). The code was modified to consider only clusters with robust statistical support (UFBoot ≥ 90%) (available at https://github.com/alex-ranieri/denvLineages/blob/main/scripts/annotate_json.py). This lineage definition method strictly considered the occurrence of, at least, one amino acid mutation within minimally sized groups (n ≥ 10) where at least 90% of sequences presented the mutation at every possible hierarchical depth level. To define a lineage, it was considered a GRI (Genotype Representation Index) of 1, implying, on average, that a single amino acid mutation differentiates the lineage from a randomly chosen sample on the phylogenetic tree.

The Autolin output underwent post-processing using a custom Python script (available at https://github.com/alex-ranieri/denvLineages/blob/main/scripts/posProcessingAutolin_aa.py) to ensure alignment with Augur Clades standards, as implemented within the Augur v24.1.0 toolkit, which became the primary platform for lineage assignment. This post-processing step involved manual curation of both the lineage designations generated by Autolin and the corresponding mutation table. The curation process focused on retaining only those mutations observed in all, or at least 90%, of sequences within a given branch. This selection strategy aimed to identify mutations with high discriminatory power, capable of reliably differentiating the corresponding phylogenetic nodes.

To ensure lineage assignments transcended serotype boundaries and adhered to established classification criteria, a systematic approach was implemented to link lineage-defining mutations with those previously associated with distinct genotypes.

This validation process leveraged a hereditary framework, where mutations associated with designated lineages were compared against the established genotype mutation table curated within the Nextstrain dengue build (https://github.com/nextstrain/dengue). This methodology was consistently applied across all hierarchical levels, encompassing both primary lineages and any identified sublineages.

Lineage labeling definitions

The proposed nomenclature system for DENV lineages follows a hierarchical structure of serotype, genotype, subgenotype, lineage, and sublineage, adapting the rationale used in the Pango system (Rambaut *et al.*, 2020). At the genotype level, a combination of numbers and letters designated both the serotype and genotype classification. For instance, "1V" means DENV-1 genotype V. The subgenotype nomenclature was chosen because it represents the initial hierarchical subdivision designated by Autolin software, following the genotype level. Thus, subgenotypes were defined using a period (".") followed by a capital letter ("A," "B," etc.) assigned sequentially. This alphabetical attribution commenced with the letter "A" for clarity and ease of reference. For instance, "1IV.A" represents a subgenotype of DENV-1 genotype IV. At the same time "1V.A" denotes a subgenotype of DENV-1 genotype V. Lineages and sublineages were labeled using ascending numeric identifiers separated by periods to indicate hierarchical depth. For clarity, this labeling system is limited to three numerical levels. When the fourth level is reached, an additional letter followed by a period and another number is appended to the lineage label. Following this logic, "2II.D.B.1" signifies a sublineage derived from lineage "2II.D.10.1.2" instead of using a longer string like "2II.D.10.1.2.1". The labeling scheme reflects the order of nodes within each genotype based on the tree topology.

Implementation of lineage assignment

To facilitate the classification of novel DENV sequences into defined lineages, a dedicated Nextclade dataset was constructed. This dataset (available at https://github.com/alex-ranieri/denvLineages/tree/main/Nextclade_V2_data) serves as a reference for lineage assignment and leverages the Nextclade framework's robust

phylogenetic inference capabilities (Aksamentov *et al.*, 2021). By incorporating this dataset within the lineage assignment workflow, newly obtained DENV sequences can be efficiently classified based on their genetic relatedness to previously defined lineages.

Results

Phylogenetic lineage-labeled tree

A total of 5,084 and 3,672 DENV-1 and DENV-2 genomes, respectively, were employed to construct the lineage-labeled phylogenetic trees. The temporal distribution of these sequences across collection years is depicted in Figure 1. Furthermore, the geographical distribution across continents and countries is presented in Figure S2. Autolin successfully assigned lineage labels to most of the DENV-1 and DENV-2 sequences. It labeled 5,036 (DENV-1) and 3,668 (DENV-2) sequences, representing approximately 99% of each dataset. This translates to a high level of representativeness for the lineage assignment process. For DENV-1, Autolin generated 192 unique annotation labels across 6 hierarchical levels, while for DENV-2, 148 unique labels were generated across 8 hierarchical levels. Details regarding the GRI values and initial labels are provided in Supplementary Files 3 and 4. The top hierarchical levels annotated by Autolin (i.e., the inner nodes of the tree) reflected the DENV genotype labels.

Following manual curation and application of the established lineage labeling definitions, a comprehensive DENV lineage classification system was developed (Figure 2). Notably, for DENV-1, a novel genotype containing 22 sequences (named 1VI) was identified, which was observed circulating in Africa with the most recent sample collected in 2021, and with no assigned subgenotypes. In total, 26 subgenotypes were designated for DENV-1, encompassing 86 distinct lineages and their sublineages. Interestingly, subgenotype designations were absent for 1II, while 1V displayed the highest subgenotype diversity within DENV-1 (n=20).

For DENV-2, 23 subgenotypes were designated, encompassing 66 lineages and their sublineages. The genotype 2V exhibited the greatest subgenotype richness within DENV-2 (n=15), whereas 2IV lacked subgenotype designations. Lineage-labeled phylogenetic trees for DENV-1 and DENV-2 are available for visualization at

https://nextstrain.org/fetch/raw.githubusercontent.com/alex-ranieri/denvLineages/main/Nextclade_V2_data/DENV1/tree.json?branchLabel=lineage&c=lineage_membership&p=full and https://nextstrain.org/fetch/raw.githubusercontent.com/alex-ranieri/denvLineages/main/Nextclade_V2_data/DENV2/tree.json?branchLabel=lineage&c=lineage_membership&p=full, respectively.

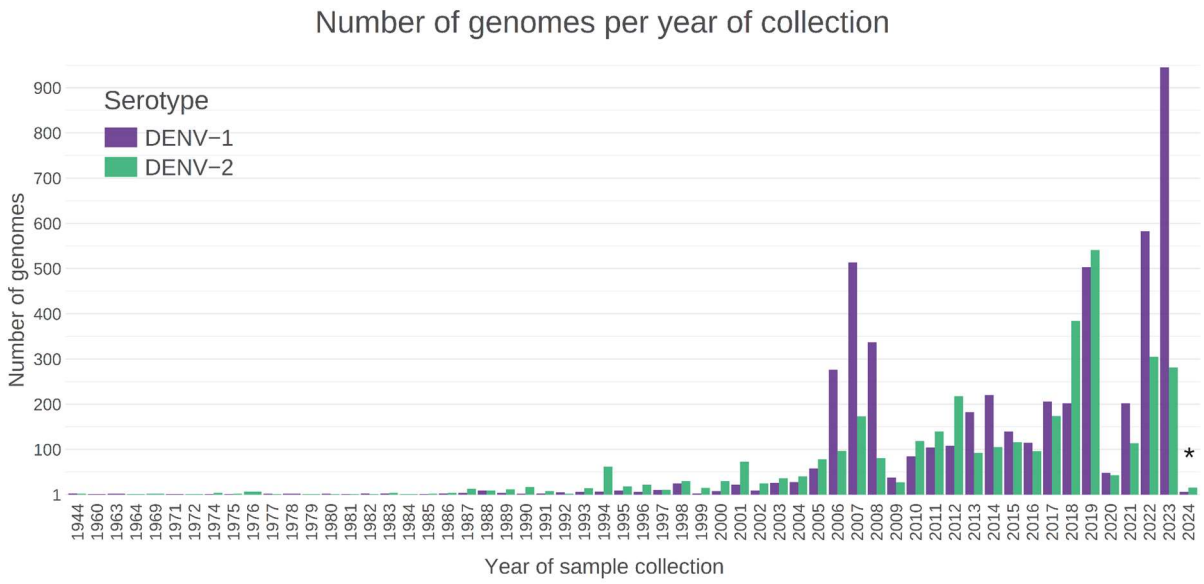

Figure 1. Temporal Distribution of DENV Genomes Employed for Lineage Assignment. The data are grouped by collection date. * For 2024, sequences were retrieved from GISAID up to January 4th for DENV-1 and January 11th for DENV-2.

In DENV-1 lineage classification, the NS5, E, and NS3 genes served as the primary determinants, being utilized in 98 (34.5%), 44 (15.5%), and 35 (12.3%) instances, respectively. Similarly, DENV-2 lineage definitions were primarily established based on mutations within the NS5, E, and NS2A genes, with these genes being employed in 64 (29.1%), 35 (15.9%), and 31 (14.1%) instances, respectively. A comprehensive mutation table, compatible with the Augur Clades, has been compiled and is available for reference at https://github.com/alex-ranieri/denvLineages/tree/main/mutation_tables.

Implementation of the lineage classification system

To promote the widespread adoption of the proposed DENV lineage classification system, a dedicated Nextclade dataset was constructed. A web version

for DENV-1 and DENV-2 are available for lineage assignment at https://v2.clades.nextstrain.org/?dataset-url=https://github.com/alex-ranieri/denvLineages/tree/main/Nextclade_V2_data/DENV1 and https://v2.clades.nextstrain.org/?dataset-url=https://github.com/alex-ranieri/denvLineages/tree/main/Nextclade_V2_data/DENV2, respectively. This dataset serves as a reference resource for assigning the lineage of novel DENV genomes generated by ongoing genomic surveillance initiatives. To evaluate the accuracy of the system in assigning lineages to novel sequences, a validation approach was employed using Nextclade v2.14.0. This involved the use of genomic sequences (complete or partial) deposited in GISAID during 2024. Forty DENV-1 sequences with collection dates ranging from January 5th to 31st and 14 DENV-2 sequences collected between January 12th and 30th were selected for analysis (data accessed on GISAID on March 20th, 2024). All obtained samples were from Brazil.

Analysis of DENV-1 sequences revealed their assignment to subgenotypes 1V.T (n=31) and 1V.J (n=8), encompassing their constituent lineages. Notably, one sequence was classified solely to genotype 1V without subgenotype designation (Figure 3 left). The genomic coverage of these sequences ranged from 74% to 97%. Both subgenotypes were first identified in 2021 and have exhibited ongoing circulation (including their associated lineages) within Brazil up to the beginning of 2024. Analysis of DENV-2 sequences revealed that a majority of sequences (n=12) were assigned to lineage 2II.D.10.1.1 (Figure 3 right). This lineage was first detected in Peru in September 2019, but its subsequent circulation has been predominantly observed in Brazil. Notably, additional detections of 2II.D.10.1.1 were identified in Paraguay and Bolivia. The most recent samples for this lineage were collected in Brazil during 2024, suggesting ongoing circulation. Lineage 2III.C.4.1 was primarily identified in Brazil, being first detected in 2016, with the most recent sequence collected in August 2023. Finally, one sequence was classified solely to genotype 2V (Asian I) without further lineage designation. The genomic coverage of DENV-2 genome sequences ranged from 74% to 99%.

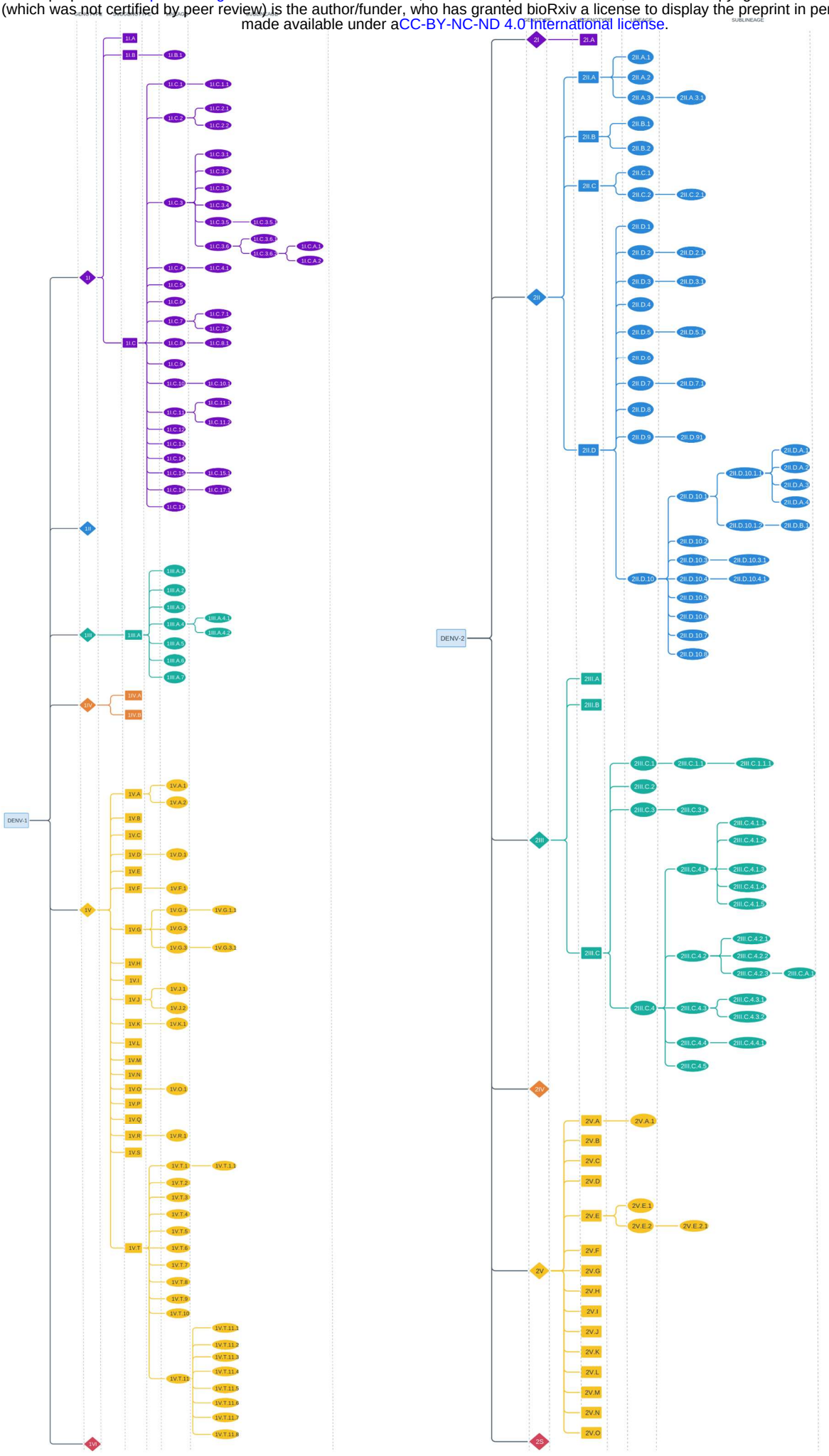

Figure 2. DENV Lineage Classification System. Schematic depicting the proposed DENV lineage classification system for serotypes 1 (left) and 2 (right). The system incorporates a hierarchical structure for each serotype, encompassing four main levels displayed from left to right: genotype, subgenotype, lineage, and sublineage.

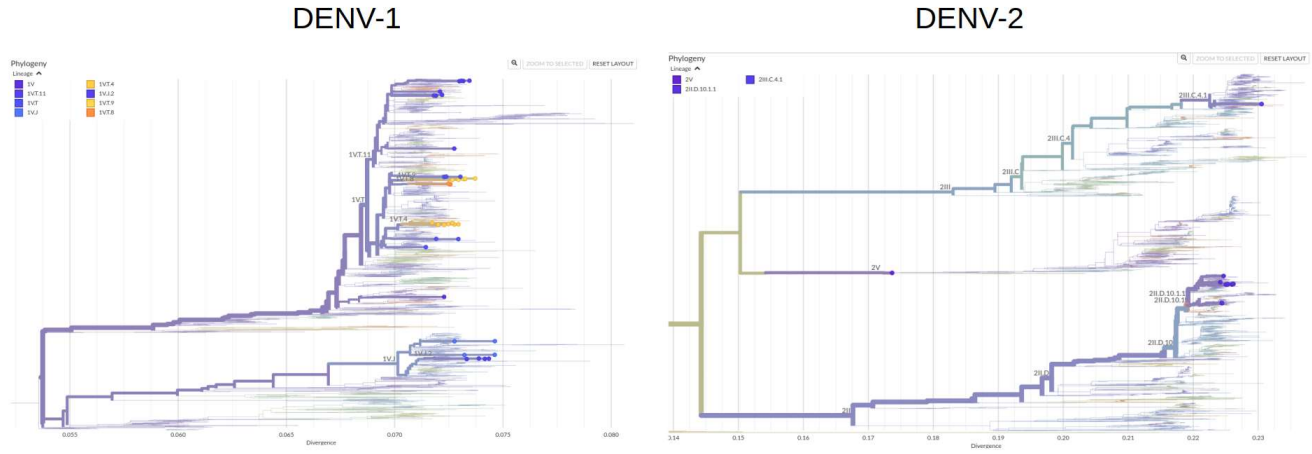

Figure 3. Genetic Diversity of Dengue Virus (DENV) Sequences in Brazil, 2024. Analysis of distinct subgenotype distributions for DENV-1 (40 sequences) is on the left panel; DENV-2 (15 sequences) sequences are displayed on the right panel. GISAID accession IDs can be found in Supplementary Files 5 and 6.

Discussion

The escalating arboviral threat posed by DENV necessitates the development of a robust, global genotyping framework. This framework would facilitate the contextualization of spatiotemporal epidemiological data, enabling a deeper understanding of DENV dynamics. Endemic persistence and incursions into previously non-endemic regions have significantly shaped DENV's genetic diversity and evolutionary divergence (Chen and Vasilakis, 2011; Fontaine *et al.*, 2018).

DENV exhibits significant genetic heterogeneity, organized hierarchically into serotypes, genotypes, and subgenotype clades. However, inconsistent nomenclature systems, a lack of standardized classification procedures, and ambiguities in reporting antigenic diversity hinder comprehensive analysis (Cuypers *et al.*, 2018). A recent study proposed a unified global genotyping framework for DENV-1, utilizing complete E gene sequences from diverse epidemic regions to address these concerns (Li *et al.*, 2022). This framework offers a standardized approach, revealing distinct global epidemic patterns, identifying persistent transmission zones and emerging outbreaks, and highlighting the necessity for a coordinated global surveillance platform to combat

DENV's expansion. However, there is a lack of lineage classification systems implemented within bioinformatic tools for DENV classification below the genotype level.

Our study addresses this gap by proposing a novel nomenclature system for DENV lineages, currently encompassing DENV-1 and DENV-2. This system aims to enhance viral surveillance and classification addressing key aspects. Firstly, it will facilitate improved monitoring efforts by providing a clear and consistent method for lineage assignment. Secondly, it will enable researchers to more precisely track the introduction and dissemination of specific regional lineages. Most importantly, our system acknowledges the growing volume of complete DENV genomes sequenced in recent years, promoting improved identification, reporting, and understanding of the virus.

DENV-1, Brazil's most prevalent and continuously circulating serotype, is traditionally classified into five genotypes (Goncalvez *et al.*, 2002). To enhance the resolution of DENV-1 classification, we propose a more granular system that categorizes this serotype into 26 subgenotypes and 86 lineages. This refined approach facilitates more precise identification and reporting of analyzed genomic samples. For instance, a study by Adelino et al., 2021, sought to characterize the diversity of DENV-1 genotype V, identifying three distinct clades (I, II, and III). The authors suggested that clades II and III replaced clade I in 2019. When analyzed using the proposed classification system, the sequences from those clades were grouped as follows: Clade I - Lineage 1V.A.1; Clade II - 1V, 1V.G.1 and 1V.G.1.1; Clade III - 1V. This finding shows that a concise lineage classification system allows for precise reporting of replacement events. It would be more informative to say that lineages 1V.G.1 and 1V.G.1.1, along with a diverging 1V branch, replaced lineage 1V.A.1. Moreover, an additional evolutionary event was observed within Clade II with the advent of the sublineage 1V.G.1.1, allowing precise tracking of genetic diversity. Finally, it shows that the detected replacement events occurred not necessarily spanning new lineages.

Intriguingly, the proposed classification system identified a putative sixth genotype for DENV-1, labeled as 1VI, and covers 22 sequences. These sequences are currently classified as part of genotype V based on the mutation scheme used by the Nextstrain dengue compilation (https://github.com/nextstrain/dengue). This discovery demonstrates the potential of the proposed system to uncover previously

unknown diversity within the DENV-1. The 1VI genotype has been circulating in Africa, and the most recent sample collected in 2021 does not have assigned subgenotypes. Identifying such potentially divergent lineages is crucial for evolutionary and epidemiological monitoring, particularly for tracking the emergence and re-emergence of strains in regions with limited genomic surveillance capabilities.

In investigations into the introduction of DENV-2 in Brazil, it was also observed that replacement events occurred within genotype III (Asian/American) of DENV-2 with the introduction of a fourth lineage. In a study conducted by De Jesus et al., 2020, the authors identified a sequence cluster naming it as Clade BR-4. Our classification method found that this cluster of sequences belongs to the same subgenotype (2III.C), specifically to the 2III.C.4 lineage. However, they could be distinguished into two different sublineages, with 17 sequences classified as 2III.C.4.1 and two classified as 2III.C.4.2. It is worth noting that 11 of the 17 sequences assigned to 2III.C.4.1 were classified into a sublineage called 2IIIC.4.1.4, adding a deeper comprehension on the cluster evolutionary dynamics.

As performed in the Clade BR-4 analysis, we applied our classification system to the samples from Amorim et al., 2024, which provides valuable insights into the genetic diversity of DENV-2 and its implications for epidemiology. In that study, the authors identified a group of samples calling it Lineage 5, belonging to genotype II (Cosmopolitan). These samples were cohesively grouped in our proposed classification as subgenotype 2II.D, specifically to sublineage 2II.D.10.1 within lineage 2II.D.10. However, two sequences present in Lineage 5 reported by Amorim et al., 2024, could be classified at a further level of depth, belonging to sublineage 2II.D.10.1.1, indicating genetic diversification events within the previously reported Lineage 5.

By investigating the genetic diversity of Clade BR-4 and Lineage 5, we inferred that each of these sequence groups is indeed internally related, as indicated by the authors, as they share the ancestries of lineages 2III.C.4 and 2II.D.10.1, respectively. However, they exhibit internal diversification events that should be considered from a genomic surveillance perspective.

The proposed unified classification system for DENV represents a significant advancement over scattered generic nomenclatures. This detailed and precise approach not only facilitates dynamic DENV monitoring, crucial for targeted vaccination strategies during outbreaks and vaccine introductions but also enables a

more agile response to changes in the epidemiological landscape. By allowing for specific lineage and sublineage monitoring, the system simplifies communication with public health authorities and facilitates effective tracking of potential viral replacements. Additionally, this standardized tool simplifies the interpretation of phylogenetic analyses across studies, resulting in a more accurate understanding of DENV evolution and dispersal, ultimately informing the development of effective control and prevention strategies.

Future perspectives

This study establishes the foundation for applying the proposed Dengue virus (DENV) lineage classification system to DENV-3 and DENV-4. This expansion will create a comprehensive framework for DENV lineage designation encompassing all four serotypes.

The system's user-friendliness is augmented by its online accessibility through the Nextclade V2 web application for DENV-1 (https://v2.clades.nextstrain.org/?dataset-url=https://github.com/alex-ranieri/denvLineages/tree/main/Nextclade_V2_data/DENV1) and DENV-2 (https://v2.clades.nextstrain.org/?dataset-url=https://github.com/alex-ranieri/denvLineages/tree/main/Nextclade_V2_data/DENV2). Alternatively, users can download datasets from https://github.com/alex-ranieri/denvLineages/tree/main for offline analysis using Nextclade CLI V2.14.0. Future integration with the Viral Identification Pipeline for Emergency Response (VIPER) assembly pipeline (https://github.com/alex-ranieri/viper) is planned. Additionally, datasets will be upgraded to ensure compatibility with Nextclade V3.

To facilitate the proposition of novel lineages, a comprehensive guideline document will be developed and hosted on https://github.com/alex-ranieri/denvLineages/tree/main. Finally, we aim to create a dedicated pipeline that leverages Nextclade, this DENV lineage classification system's datasets, and Autolin software to suggest new lineages for newly submitted sequences. This will further enhance DENV genomic surveillance efforts.

Acknowledgements

Funding for this work was provided by Fundação Butantan and the São Paulo Research Foundation (FAPESP) through grant number 21/11944-6, titled "Continuous

improvement of vaccines: Center for Viral Surveillance and Serological Assessment (CeVIVAS)". We express our sincere gratitude to all researchers who deposited their sequences on GISAID, particularly those affiliated with the Central Public Health Laboratories (LACEN) from the Brazilian States of Alagoas, Pará, and Paraná. We are also grateful to São Paulo City Hall. These partners are instrumental to the success of the CeVIVAS project.

References

Adams, B. *et al.* (2006) 'Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok', *Proceedings of the National Academy of Sciences of the United States of America*, 103(38), pp. 14234–14239. Available at: https://doi.org/10.1073/pnas.0602768103.

Adelino, T.É.R. *et al.* (2021) 'Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil', *Nature Communications*, 12(1), p. 2296. Available at: https://doi.org/10.1038/s41467-021-22607-0.

Aksamentov, I. *et al.* (2021) 'Nextclade: clade assignment, mutation calling and quality control for viral genomes', *Journal of Open Source Software*, 6(67), p. 3773. Available at: https://doi.org/10.21105/joss.03773.

Amorim, M.T. *et al.* (2024) 'Detection of a Multiple Circulation Event of Dengue Virus 2 Strains in the Northern Region of Brazil', *Tropical Medicine and Infectious Disease*, 9(1), p. 17. Available at: https://doi.org/10.3390/tropicalmed9010017.

Balakrishnan, V.S. (2022) 'WHO launches global initiative for arboviral diseases', *The Lancet Microbe*, 3(6), p. e407. Available at: https://doi.org/10.1016/S2666-5247(22)00130-6.

Campbell, K. *et al.* (2022) 'Making genomic surveillance deliver: A lineage classification and nomenclature system to inform rabies elimination', *PLOS Pathogens*. Edited by A.S. Lauring, 18(5), p. e1010023. Available at: https://doi.org/10.1371/journal.ppat.1010023.

Chen, R. and Vasilakis, N. (2011) 'Dengue — Quo tu et quo vadis?', *Viruses*, 3(9), pp. 1562–1608. Available at: https://doi.org/10.3390/v3091562.

Cuypers, L. *et al.* (2018) 'Time to Harmonize Dengue Nomenclature and Classification', *Viruses*, 10(10), p. 569. Available at: https://doi.org/10.3390/v10100569.

Fontaine, A. *et al.* (2018) 'Epidemiological significance of dengue virus genetic variation in mosquito infection dynamics', *PLOS Pathogens*. Edited by N.M. Ferguson, 14(7), p. e1007187. Available at: https://doi.org/10.1371/journal.ppat.1007187.

Goncalvez, A.P. *et al.* (2002) 'Diversity and Evolution of the Envelope Gene of Dengue

Virus Type 1', *Virology*, 303(1), pp. 110–119. Available at: https://doi.org/10.1006/viro.2002.1686.

Goya, S. *et al.* (2024) 'The unified proposal for classification of human respiratory syncytial virus below the subgroup level', *MedRxiv* [Preprint]. Available at: https://doi.org/https://doi.org/10.1101/2024.02.13.24302237.

Grubaugh, N.D. (2022) 'Translating virus evolution into epidemiology', *Cell Host & Microbe*, 30(4), pp. 444–448. Available at: https://doi.org/10.1016/j.chom.2022.03.006.

Ha, A.D. and Aylward, F.O. (2024) 'Automated classification of giant virus genomes using a random forest model built on trademark protein families', *npj Viruses*, 2(1), p. 9. Available at: https://doi.org/10.1038/s44298-024-00021-9.

Happi, C. *et al.* (2022) 'Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus', *PLOS Biology*, 20(8), p. e3001769. Available at: https://doi.org/10.1371/journal.pbio.3001769.

Huddleston, J. *et al.* (2021) 'Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens', *Journal of Open Source Software*, 6(57), p. 2906. Available at: https://doi.org/10.21105/joss.02906.

de Jesus, J.G. *et al.* (2020) 'Genomic detection of a virus lineage replacement event of dengue virus serotype 2 in Brazil, 2019', *Memórias do Instituto Oswaldo Cruz*, 115. Available at: https://doi.org/10.1590/0074-02760190423.

Kalyaanamoorthy, S. *et al.* (2017) 'ModelFinder: fast model selection for accurate phylogenetic estimates', *Nature Methods*, 14(6), pp. 587–589. Available at: https://doi.org/10.1038/nmeth.4285.

Katoh, K. and Standley, D.M. (2013) 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', *Molecular Biology and Evolution*, 30(4), pp. 772–780. Available at: https://doi.org/10.1093/molbev/mst010.

Li, Liqiang *et al.* (2022) 'A unified global genotyping framework of dengue virus serotype-1 for a stratified coordinated surveillance strategy of dengue epidemics', *Infectious Diseases of Poverty*, 11(1), p. 107. Available at: https://doi.org/10.1186/s40249-022-01024-5.

Lourenço, J. *et al.* (2018) 'Challenges in dengue research: A computational perspective', *Evolutionary Applications*, 11(4), pp. 516–533. Available at: https://doi.org/10.1111/eva.12554.

McBroome, J. *et al.* (2024) 'A framework for automated scalable designation of viral pathogen lineages from genomic data', *Nature Microbiology*, 9(2), pp. 550–560. Available at: https://doi.org/10.1038/s41564-023-01587-5.

Minh, B.Q. *et al.* (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*. Edited by E. Teeling, 37(5), pp. 1530–1534. Available at: https://doi.org/10.1093/molbev/msaa015.

512 Mustafa, M.S. *et al.* (2015) 'Discovery of fifth serotype of dengue virus (DENV-5): A
513 new public health dilemma in dengue control', *Medical Journal Armed Forces India*,
514 71(1), pp. 67–70. Available at: https://doi.org/10.1016/j.mjafi.2014.09.011.

515 Nanaware, N. *et al.* (2021) 'Dengue Virus Infection: A Tale of Viral Exploitations and
516 Host Responses', *Viruses*, 13(10), p. 1967. Available at:
517 https://doi.org/10.3390/v13101967.

518 Nunes, M.R.T. *et al.* (2014) 'Air Travel Is Associated with Intracontinental Spread of
519 Dengue Virus Serotypes 1-3 in Brazil', *PLoS Neglected Tropical Diseases*, 8(4).
520 Available at: https://doi.org/10.1371/journal.pntd.0002769.

521 Oude Munnink, B.B. and Koopmans, M. (2023) 'Tracking SARS-CoV-2 variants and
522 resources', *Nature Methods*, 20(4), pp. 489–490. Available at:
523 https://doi.org/10.1038/s41592-023-01833-y.

524 PAHO (2024) *Situation Report No 9 - Dengue Epidemiological Situation in the Region
525 of the Americas - Epidemiological Week 08, 2024*. Available at:
526 https://www.paho.org/en/documents/situation-report-no-3-dengue-epidemiological-
527 situation-region-americas-epidemiological (Accessed: 21 March 2024).

528 PAHO and WHO (2024) *Epidemiological Alert: Dengue in the Region of the Americas.
529 16 February 2024*. Available at:
530 https://www.paho.org/en/file/139834/download?token=x7ngARBs (Accessed: 15
531 March 2024).

532 Rambaut, A. *et al.* (2020) 'A dynamic nomenclature proposal for SARS-CoV-2
533 lineages to assist genomic epidemiology', *Nature Microbiology*, 5(11), pp. 1403–1407.
534 Available at: https://doi.org/10.1038/s41564-020-0770-5.

535 Rico-Hesse, R. (1990) 'Molecular evolution and distribution of dengue viruses type 1
536 and 2 in nature', *Virology*, 174(2), pp. 479–493. Available at:
537 https://doi.org/10.1016/0042-6822(90)90102-W.

538 Sagulenko, P., Puller, V. and Neher, R.A. (2018) 'TreeTime: Maximum-likelihood
539 phylodynamic analysis', *Virus Evolution*, 4(1). Available at:
540 https://doi.org/10.1093/ve/vex042.

541 Wallau, G.L. *et al.* (2023) 'Arbovirus researchers unite: expanding genomic
542 surveillance for an urgent global need', *The Lancet Global Health*, 11(10), pp. e1501–
543 e1502. Available at: https://doi.org/10.1016/S2214-109X(23)00325-X.

544 Wang, W.-H. *et al.* (2020) 'Dengue hemorrhagic fever – A systemic literature review
545 of current perspectives on pathogenesis, prevention and control', *Journal of
546 Microbiology, Immunology and Infection*, 53(6), pp. 963–978. Available at:
547 https://doi.org/10.1016/j.jmii.2020.03.007.

548 WHO (2023) *Dengue – Global situation*, *Disease Outbreak News*. Available at:
549 https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498
550 (Accessed: 14 March 2024).

551 WHO (2024) *Tracking SARS-CoV-2 variants*. Available at:

https://www.who.int/activities/tracking-SARS-CoV-2-variants (Accessed: 15 March 2024).

Supplementary Figures

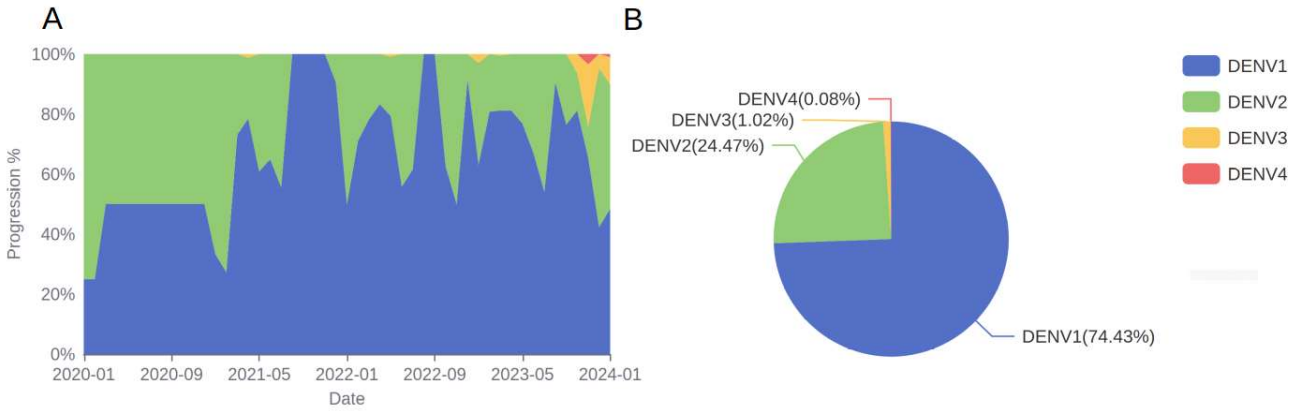

Figure S1. Distribution of DENV Genome Sequences across Serotypes in Brazil. A. Percentage distribution of DENV genome sequences by serotype, spanning from January 2020 to January 2024. **B.** Serotype prevalence of DENV based on genome sequencing data. Data sourced from GISAID as of March 15, 2024, encompassing 2,693 genome sequences.

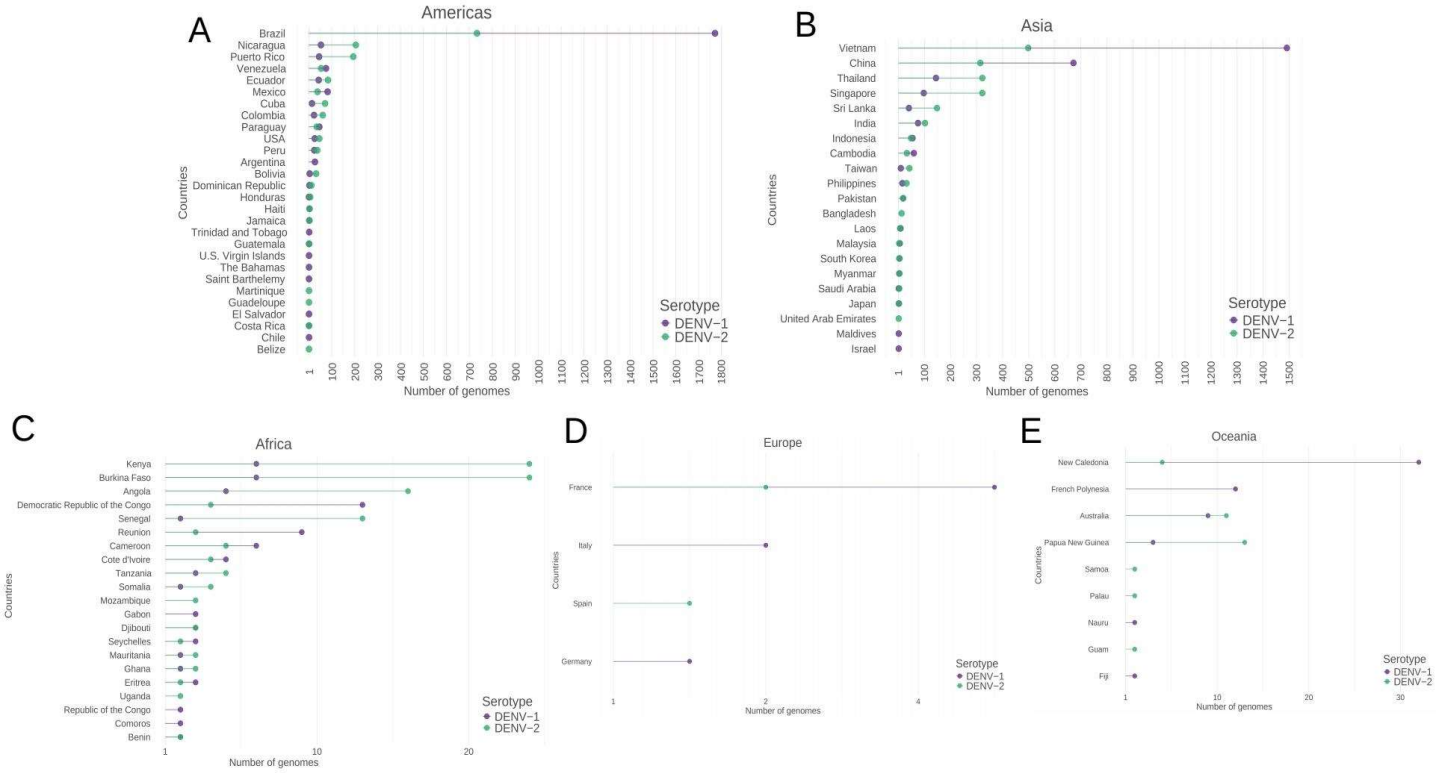

Figure S2. Geographical Distribution of DENV Genomes Employed for Lineage Assignment. The data is presented according to the continent and country of origin for each sequence.