

Genomic and structural features of the Yellow Fever virus from the 2016-2017 Brazilian outbreak

Mariela Martínez Gómez¹, Filipe Vieira Santos de Abreu^{2,3}, Alexandre Araujo Cunha dos Santos¹, Iasmim Silva de Mello¹, Marta Pereira Santos¹, Ieda Pereira Ribeiro¹, Anielly Ferreira-de-Brito², Rafaella Moraes de Miranda², Marcia Gonçalves de Castro², Mario Sergio Ribeiro⁴, Roberto da Costa Laterrière Junior⁵, Shirlei Ferreira Aguiar⁶, Guilherme Louzada Silva Meira⁶, Deborah Antunes⁷, Pedro Henrique Monteiro Torres⁷, Ana Carolina Paulo Vicente⁸, Ana Carolina Ramos Guimarães⁹, Ernesto Raul Caffarena⁷, Gonzalo Bello¹⁰, Ricardo Lourenço-de-Oliveira^{2*}, Myrna Cristina Bonaldo^{1*}

1- Laboratório de Biologia Molecular de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.

2- Laboratório de Mosquitos Transmissores de Hematozoários, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.

3- Instituto Federal do Norte de Minas Gerais, Salinas, MG, Brazil.

4- Superintendência de Vigilância Epidemiológica e Ambiental, Secretaria Estadual de Saúde, Rio de Janeiro, Brazil.

5- Núcleo Especial de Vigilância Ambiental, Secretaria Estadual de Saúde do Espírito Santo, Vitória, Brazil.

6- Laboratório Central de Saúde Pública Noel Nutels (LACEN-RJ), Rio de Janeiro, Brazil.

7- Programa de Computação Científica (PROCC), Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.

8- Laboratório de Genética Molecular de Microorganismos, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.

9- Laboratório de Genômica Funcional e Bioinformática, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.

10- Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil.

31

32 ***Corresponding author:**

33 Correspondence and requests for materials should be addressed to R.L.-d.-O. (email:

34 lourenco@ioc.fiocruz) or M.C.B. (mbonaldo@ioc.fiocruz.br).

35

36 Keywords: Yellow fever virus, 2016-2017 Brazilian outbreak, amino acid changes,
37 genetic diversity, evolution

38

39

40

41

42

43

44

45

46

47

ABSTRACT

Brazil has been suffering a severe sylvatic epidemic of yellow fever virus (YFV) since late 2016. Analysis of full-length YFV genomes from all hosts involved in the Brazilian 2017 outbreak reveals that they belong to sub-lineage 1E within modern-lineage, but display several unique amino acid substitutions in highly conserved positions at NS3 and NS5 viral proteins. Evolutionary analyses indicate that YFV carrying that set of amino acid substitution circulates in the Southern Brazilian region for several months before being detected in December 2016. Structural and selection analyses support that some of these substitutions were under positive selection and could impact enzyme structure and function. Altogether, this evidence demonstrated that the current Brazilian YFV carries unique amino acid signatures in the non-structural proteins and support the hypothesis that those substitutions may be affecting the viral fitness and transmissibility.

INTRODUCTION

Yellow fever (YF) is a viral disease transmitted by the bite of infected mosquitoes in Africa and South America, affecting around 200,000 people annually, mostly in Africa (1-3). There are two main epidemiological cycles: the enzootic sylvatic cycle where the virus is transmitted between non-human primates (NHP) and wild arboreal mosquitoes of genus *Aedes*, *Haemagogus* and *Sabethes*, and in which humans can be accidentally infected, and the urban cycle where inter-human transmission is ensured by the domestic and anthropophilic mosquito *Aedes aegypti* (3). While only the sylvatic cycle has been reported in the Americas during nearly the last 75 years, in Africa people may acquire the infection in both cycles, besides in an intermediate cycle occurring in rural areas close to forests (2, 4).

The causative agent is the yellow fever virus (YFV) (genus: *Flavivirus*, family: *Flaviviridae*), presenting a single-positive-sense RNA genome, containing a 5' end cap structure, that is translated in a single immature polyprotein precursor. The precursor polyprotein is cleaved into three structural proteins, capsid (C), envelope (E), and membrane protein (M) and seven non-structural proteins, NS1, NS2A, NS2B, NS3, NS4A NS4B, and NS5 (2). The virus was originated in Africa, where five genotypes have been documented, being two from West Africa (West Africa I and II) and three in East and Central Africa (East Africa, East/Central Africa, and Angola). The YFV virus has spread from Africa to the Americas together with the invasive mosquito *Ae. aegypti* where it evolved in the last four centuries into two genotypes (South America I and II) derived from the Western African ancestor (5-7). The South American genotype I is the most spread and frequently detected during the epizootics and epidemics waves in Brazil and other countries of South America (8, 9). Until the 1990's, the transmission area in Brazil was primarily limited to the Amazon forest, in the Northern, and the savanna-like *cerrado*, in Center-West region. However, in about two decades, the YFV territory has progressively expanded Southward and Eastward reaching the Atlantic forest and other biomes from the country's most populated regions(10). During this boundary expansion, five viral sub-lineages (1A to 1E) successively arose within the genotype I. They were distinguished by analysis of partial nucleotide sequencing of the YFV genome particularly the pre-membrane and envelope (prM/E) gene junction (8, 9, 11). Most recently, South America genotype I was divided into two major lineages named as Old lineages (enclosing Old Para, and 1A, 1B, and 1C sub-lineages) and Modern lineage (enclosing Trinidad and Tobago, and 1D and 1E sub-lineages) (11, 12).

A rapid expansion of the YFV area has reported since late 2016 in Southeast Brazil (Fig.1) (13, 14). From December 2016 to May 2017, the YFV quickly spread from the

transition zone between the *cerrado* and the Atlantic forest inland of Minas Gerais State (MG) to the coastal areas in the Espírito Santo (ES) and Rio de Janeiro (RJ) states, then approaching to densely populated sites with insignificant vaccination coverage under influence of the rain forest. A YFV outbreak of unprecedented sanitary severity and causing an ecological disaster was recorded. In a few months, around 3,850 NHPs died, and nearly 800 human cases with 435 deaths were registered, of which 274 were YF confirmed until July 2017 (13, 14). Interestingly, during this ongoing outbreak analysis of complete nucleotide genome sequences of the YFV obtained from the blood of two howler monkeys from a single locality in ES confirmed that they cluster within the sub-lineage 1E. Furthermore, these strains revealed new polymorphisms comprising several amino acid substitutions mainly located in the components of the viral replicase complex, in the protease domain of the NS3 protein, and in the methyltransferase (MTase) and RNA-dependent RNA polymerase (RdRp) domain of the NS5 protein(15). The NS3 protein, a viral multi-functional protein, carries a chymotrypsin-like serine protease activity (NS3pro) at its N-terminal and an helicase activity (NS3hel) at its C-terminal domain(16). The NS3pro associated with its cofactor NS2B cleaves all cytoplasmic protein junction sites of the precursor polyprotein. The NS5 protein contains two functional domains with a capping related MTase and the central replication enzyme RdRp (17-19). Connecting MTase and RdRp, there is a linker of 5–6 residues (residues 266–271; NS5 Dengue virus position), which has an essential role in NS5 conformation and protein activity (20). Also, NS3 and NS5 have been associated with innate immune system evasion (21, 22). It was unclear, however, whether the observed amino acid substitutions are genetic signatures of the most recent YF outbreaks as few complete genomes of current circulating viral strains were available (8, 15).

In this study, we elucidated the complete genome sequence of 12 YFV strains from three hosts (NHPs, mosquitoes, and humans) involved in the transmission cycle of the current Brazilian outbreak in two Southeastern states (RJ and ES). Sequences were analyzed to establish whether specific amino acid changes previously observed are fixed in other recent YFV samples and therefore constitute a molecular signature of the 2017 YFV. We also created homology models for NS3 and NS5 to determine the location and potential effects of amino acid substitutions on NS3pro and NS5 proteins. Moreover, we performed phylogenetic and evolutionary studies to analyze the codons that might be under positive selection pressure and to estimate the time of the most recent common ancestor (T_{MRCA}) of Brazilian YFV 2017 samples.

RESULTS

YFV samples geographic distribution and genome characterization

From February to June 2017, we collected 15 YFV samples from five distinct infected host species including mosquitoes, NHP and humans from 11 localities belonging to three main river basins across the current epidemic/epizootic territory in the Atlantic coast in ES and RJ (Table 1; Fig. 1). We determined and analyzed the whole YFV genome of twelve samples: two pools of infected mosquitoes belonging to two species *Haemagogus leucocelaenus* and *Hg. janthinomys*, six from NHP (four howler monkeys and two marmosets), and four from human cases (Table 1). Partial sequences of the YFV genome were also obtained from an additional human case (H189) (data not shown).

The comparison of all the YFV genomes from the ongoing Southeast Brazilian outbreak reveals low genetic variation, providing a mean nucleotide identity of 99.8 % and amino acid identity values ranged from 99.9% to 100%. However, part of the

nucleotide variations is non-synonymous, leading to new amino acid substitutions in the polyprotein sequence (Fig. 2; Supplementary file 1). Regardless the host, all the 2017 YFV Brazilian genomes display a set of eight unique amino acid substitutions, that we have recently identified in two YFV samples from infected howler monkeys (ES-504 and ES-505) from ES state (15). Remarkably, the comparison with the genome of other South America strains confirmed that these polymorphisms are only present in the Brazilian strains from the ongoing outbreak when comparing with other South American strains. They localize at C protein (V108I), at NS3pro (E1572D; R1605K), at NS5 in MTase domain (K2607R; V2644I; G2679S), and at NS5 in RdRp domain (V3149A; N3215S) (Fig. 2). Nevertheless, the partial nucleotide sequence from H189 also displayed all amino acid changes detected in the other sequences, except for the mutation in the C protein. Moreover, all 2017 YFV Brazilian strains also share an amino acid change at position N/D2803S that was previously observed only in a Venezuelan strain isolated in 2006 (GenBank, KM388818). We also detected additional amino acid substitutions, which are not present in all the 2017 Brazilian genomes (Fig. 2). Accordingly, the YFV H199 sequence shows a change from an alanine (A) to a serine (S) at the amino acid position 826, corresponding to the NS1 protein (position 48). The YFV genomes RJ95, RJ96, RJ97, H191 and PA193, show a substitution from an isoleucine (I) to a valine (V) at position 2176 (NS4A, position 77). The phylogenetic analysis of prM/E sequences indicates that all 2017 YFV Brazilian strains cluster inside sub-lineage 1E of the Modern lineage of South America genotype I in a monophyletic clade with high support (bootstrap = 87 %) (Fig. 2 – Fig. Supplement 1). The same clustering pattern was obtained when we performed the phylogenetic analyses of either NS3 or NS5 nucleotide sequences (Fig. 2 – Fig. Supplement 2).

Modeling and Structural Analysis of NS2B/NS3 and NS5 proteins

We created structural models of NS3 and NS5 proteins to gain some insights into the structural and functional effects of these amino acid substitutions. Initially, prior to the structural model generation, we aligned the NS3 and NS5 amino acid sequences of the prototype 2017 Brazilian YF virus (strain ES505) (15) with the different template sequences (Fig. 3– Fig. Supplements 1; Fig. 4– Fig. Supplements 1).

The effect of the amino acid substitutions on both NS3 and NS5 proteins was evaluated through hydrogen bond formation and electrostatic analysis. The two substitutions found in NS3, E88D and R121K (polyprotein position: E1572D; R1605K, respectively) are conservative and, as such, they have little impact on the surface electrostatic potential (Fig. 3 C, D). For the E88D substitution, a small structural change was observed, which mainly consisted of lysine 174 (polyprotein position: K1658) side chain displacement due to the loss of a hydrogen bond with the protein backbone (Fig. 3 E, F). On the other hand, the R121K substitution is located near the NS2B binding groove and might influence the interaction between these two molecules. Hydrogen bond analysis indicates that K 121 could favor the formation of a hydrogen bond with threonine 77 of NS2B (polyprotein position: T1431), whereas such an interaction was not identified in 2010 model (Fig. 3 A, B). This interaction could, in turn, modulate the NS3-NS2B binding affinity and, thus, the protease efficiency.

The three first amino acid substitutions in NS5 are clustered in the MTase domain, whereas the remaining ones are in the RdRp domain. All amino acid substitutions at the MTase domain are conservative, but they are spatially adjacent. Arginine 101 (polyprotein position: R2607) alpha carbon is 9.7 Å away from that of isoleucine 138 (polyprotein position: I2644), which in turn is 10.6 Å away from serine 173 alpha carbon (polyprotein position: S2679). These three residues face the RdRp domain tunnel opening (Fig. 4B), which presents a basic electrostatic profile to accommodate the YFV RNA

molecule (Fig. 4 – Fig. Supplement 2), suggesting that they may influence the enzyme activity.

Additionally, the N297S substitution (polyprotein position: N/D/S 2803) from the RdRp domain, although being conservative, is located near the hinge domain. The remaining two amino acid alterations: V643A (polyprotein position: V/A 3149) and N709S (polyprotein position: N/S 3215) are located at the protein surface and are also conservative. The combined effect of the mutations on the NS5 protein dynamics was assessed through molecular dynamics simulations, which indicate that the NS5 protein from 2017 sample has a decrease in fluctuation around the hinge region (Fig. 4– Fig. Supplement 3).

Selection analyses

The mean dN/dS ratio of substitutions per site estimated by the SLAC method for the South America genotype I (SAI) and West Africa (WA) data sets was 0.05 and 0.04, respectively; thus indicating that purifying selection was the main evolutionary force of both YFV genotypes. Tests for negative/positive selection, however, reveal some important differences in the evolutionary dynamics of both genotypes (Fig. 5).

Selection analysis of the SAI dataset identifies 13 codons with evidence of positive selection, including nine sites with evidence of episodic diversifying selection (MEME), four sites with evidence of pervasive diversifying selection (FEL and/or FUBAR) and one site identified by all three algorithms (FEL, MEME, and FUBAR). Most sites (69%, 9/13) under positive selection were concentrated in a short genome segment (2,100-2,850 codon positions) coding for non-structural proteins NS4A (I2176V), NS4B (H2311L, N2408S, K2502N, T2503I) and NS5 (R2601K, R2640P, D2647V, N/D2803S) (Fig. 5 – Fig. Supplement 1). Selection analysis of the WA dataset, by contrast, detected no sites under pervasive diversifying selection and identified 18 sites

with evidence of episodic diversifying selection (MEME) uniformly distributed along structural and non-structural proteins (Fig. 5 – Fig. Supplement 1).

Evolutionary analyses

The Bayesian MCC tree inferred from the complete coding sequence (CDS) of YFV South American genotypes I and II reveals that YFV strains from the ongoing Southeast Brazilian outbreak grouped in a highly supported (*Posterior Probability* [*PP*] = 1) monophyletic cluster nested within sub-lineage E strains of the modern-lineage (Fig. 6). We further observed that 2017 YFV Brazilian strains segregate in two reciprocally monophyletic subgroups, one sub-cluster comprising strains of mosquitoes, NHP and humans from RJ and ES states (sub-clade A) (*PP* = 0.29), and one highly supported (*PP* = 0.99) sub-cluster containing YFV strains of NHP and humans from the state of RJ (sub-clade B). The median substitution rate of YFV South American genotypes complete genomes was estimated at 3.5×10^{-4} subst./site/year (95% HPD: $2.4-4.8 \times 10^{-4}$ subst./site/year), in agreement with previous estimations(12). The median T_{MRCA} for all Brazilian 2017 YFV strains was estimated in April 2016 (95% HPD: July 2015 - October 2016) and for the sub-clade B at October 2016 (95% HPD: April 2016 - January 2017). In addition, the median T_{MRCA} of Venezuelan and Brazilian 2017 YFV strains was estimated as occurred 24 years ago.

DISCUSSION

In the 2016-2017 YFV outbreak in Southeast Brazil, most of the epizootics and human cases originally occurred in inland rural areas of MG and subsequently in Eastern Atlantic coastal sites under the influence of the ES rain forest segments. After that, the YFV spreading approached and even touched the Great Metropolitan areas of Vitoria (ES) and Rio de Janeiro (RJ), where lived nearly 1.8 and 12.3 million unvaccinated

inhabitants, respectively. Alarming, these densely populated areas host some of the busiest South American airports, ports and road networks, are highly infested by YFV competent urban vectors, *Aedes aegypti* and *Ae.albopictus*, and have repeatedly been the territory of severe dengue epidemics. All together, these ecological and sanitary marks raised concern about the potential risk of YFV to reemerge in an urban cycle in Brazil as well as to spread to other countries and continents rapidly(23, 24). Furthermore, YFV strains characterized by a set of amino acid polymorphism were identified during this outbreak, and their biological impact needed to be investigated (15).

Regardless of deriving from five distinct host species infected in a six month lag in 11 sites dispersed along 350 Km of the Southeast Brazilian coast across the outbreak territory including the Great Metropolitan area of Rio de Janeiro, the YFV samples analyzed in the current study are almost identical at the nucleotide and amino acid levels. Also, they share a molecular signature represented by nine amino acids, being eight in highly conserved positions at NS3 and NS5 encoding regions, and one in the structural capsid protein. Previous analysis of 2017 YFV from two howler monkeys from a single site in ES had not considered the substitution at position 2803 at NS5. However, in the current study, the analysis of a higher number of samples allowed the identification of one more substitution at position 2803 at NS5 in all 2017 YFV. This molecular signature represented by nine amino acids have never been observed before in South-American and African genotypes (15).

Hypothetically, amino acid changes at conserved protein positions of NS3 and NS5 may have a role in the capacity of viral infection to vertebrate and invertebrate hosts and thus accelerating the spreading of the ongoing outbreak. The NS3 and NS5 proteins have multiple enzymatic activities essential for viral RNA replication and 5'-capping (25-27). For these reasons, we also investigated the impact of the identified amino acid

substitutions through the structural analysis of NS3 and NS5 protein models. Even though all the amino acid substitutions were mainly conservative, they occur close to domains that might be affected by these subtle modifications. Furthermore, we were able to detect features that may correlate with an increase in enzymatic efficiency and constitute an advantage in viral dissemination. In the NS3 protein, the R121K substitution is located in the region responsible for the interaction with NS2B, the cofactor for the proteolytic activity of this enzyme. Although both residues bear a positive charge, lysine was shown to potentially establish a hydrogen bond with NS2B, due to the less bulky side chain. For the NS5 protein, we found that the three amino acid substitutions located at the MTase domain were spatially contiguous and could influence the relative orientation between the two domains. The N297S substitution might also have a significant role in the enzymatic efficiency since is located near the hinge domain. Hence, molecular dynamics simulations have shown that all these substitutions combined may have a stabilizing effect on the linker domain, which has been demonstrated to influence the enzyme processivity directly and viral replication in Dengue 4 virus *in vitro* models (28, 29). These findings shall be addressed in future studies considering the unrevealed diversity of the YFV in the cell culture and animal models.

A hot spot region of episodic or pervasive positive selection was identified in between codons 2100-2850 of CDS region in YFV genomes of South American genotype I, that was not detected in the West Africa genotype. Interestingly, most of these positively selected sites localize in the NS4B, and NS5 coding regions have also been described in Zika virus from the current epidemics(30). Several studies have demonstrated the role of non-structural proteins in the host innate immune response against flavivirus infection (22, 31-33). Onward, these proteins widely interact with other viral proteins and host molecules (27, 33-35). It is also interesting to note that the 2803

position, part of the molecular signature of 2017 YFV, is one of the positively selected sites. Also, two other positions positively selected (826 and 2176), presented variability in the 2017 YFV. These observations support some singularities in the evolutionary dynamics of YFV South American genotype I and also indicate that fixation of some amino acid substitutions in Brazilian 2017 YFV strains might have been driven by positive selection.

Since the beginning of the XXI, a striking spreading of the YFV has been occurring in Brazil. The outbreaks formerly constrained to the endemic Amazon and Central-Western regions have reached the South and Southeast of Brazil, where vaccination coverage against YF was insignificant until the explosion of the ongoing outbreak. It has been recently proposed that a YFV strain from Trinidad-Tobago introduced in Brazil and Venezuela in the late 1980s originate all modern-lineages strains belonging to sub-lineages 1D and 1E (12). The main source of variability related to the ancestral lineages is the occurrence of several amino acid substitutions particularly within non-structural viral proteins (12), as observed in the 2017 Brazilian outbreak YFV. All Brazilian 2017 YFV belonged to the sub-lineage 1E and clustered with the Venezuelan strains isolated in the late 2000s, consistent with the notion that ancestral YFV strain responsible for the ongoing Brazilian outbreak would have originated in Venezuela (12).

Although Brazilian 2017 YFV strains are most closely related with Venezuelan 2004-2010 YFV strains, they share a relatively distant common ancestor traced to 1993 (95% HPD: 1987-1997) (Fig. 6). This indicates that the virus may have circulated in endemic South American regions for a long period before being introduced in Southeast Brazil, but the precise pathway of viral dissemination is difficult to elucidate because the paucity of Brazilian YFV sequences sampled from endemic regions over the last 10-15 years. We estimate the median T_{MRCA} for the Brazilian 2017 YFV strains at early 2016,

suggesting that the virus circulated for several months in the Southeastern region before the ongoing outbreak was first recognized at December 2016. This pre-detection period of cryptic transmission of YFV in Southeastern Brazil is comparable to that recently estimated for Zika virus in the Northeastern Brazilian region (36).

Bayesian analysis also showed that Brazilian 2017 YFV strains segregate into two sub-clusters, one of them (sub-clade A) containing sequences from both RJ and ES, and the other (sub-clade B) comprising only sequences from RJ whose T_{MRCA} was traced to October 2016. It is also interesting to note that all YFV sequences of RJ that branched within sub-clade A were sampled from sites situated in the Paraíba do Sul basin whose tributaries born on the northern side of the Serra do Mar, a 1,500km long system of mountain ranges and escarpments that runs parallel to the Atlantic Ocean coast in Southeastern Brazil (Fig. 1). By contrast, YFV sequences of RJ that branched within sub-clade B were obtained from locations situated along the Macaé conjugated river basin whose tributaries born on the escarpments of the coastal side of the Serra do Mar. The only exception is the sample H190, which despite originating in the Paraíba do Sul river basin, it clustered in the sub-clade B. Intriguingly, the sampling location of H190 (São Fidelis) is located in the largest discontinuity of the long mountain ranges of the Serra do Mar and this YFV strain branched basally to all other sub-clade B strains. Together, these results support the occurrence of multiple independent introductions of YFV in RJ (probably from both ES and MG) and further indicate at least two main routes of viral dissemination within the state running at the northern and coastal sides of the Serra do Mar. These analyses also point that YFV dissemination through the coastal route in Rio de Janeiro probably started in São Fidelis around late 2016, but more YFV sequences from other Southeastern states are necessary to confirm this conclusion.

Future analysis based on reverse genetic approaches can contribute to establishing the role of amino acid substitutions present in Brazilian 2017 YFV in the viral fitness and transmissibility. It will also be crucial to improving the number of YFV genomes from Brazilian endemic and non-endemic regions in the last years to better understand the epidemiology during recent years.

MATERIAL AND METHODS

Ethics Statements.

This study was reviewed and approved by the Ethics Committee for human research at the Instituto Oswaldo Cruz (IOC) (CAAE 69206217.8.0000.5248), which exempted the need of a specific written informed consent from patients or their legal representatives. The protocols for mosquito rearing as well as handling and blood sampling of NHP were approved by the Institutional Ethics Committee of Animal use at IOC (CEUA licenses LW-34/2014 and L037/2016, respectively). Capture of wild NHPs and mosquitoes were approved by the Brazilian environmental authorities: SISBIO-MMA licenses 54707-137362-2 and 52472-1, and INEA license 012/2016012/2016. No specific permits are needed for conducting mosquito collection in the urban and suburban areas in Southeastern Brazil. This study does not include endangered or protected species.

Mosquito samples

Adult mosquitoes were collected with BG-sentinel adult traps baited with dry ice as a source of CO₂ as well as with an insect net when approaching to humans in the forest, at the forest fringe, and in the modified environment. The insects were immediately frozen in dry ice or N₂, transported to the laboratory, identified to species according to Consoli and Lourenço-de-Oliveira (36), and pooled according to species, the number of captured mosquitoes and collecting site and day. Entire bodies of pooled mosquitoes were ground

and treated in Leibovitz L15 medium (Invitrogen) supplemented with 4% Fetal Bovine Serum (FBS), and submitted to the RNA extraction as described elsewhere(37).

Non-human primate samples

Blood samples were taken from the femoral vein of dying NHPs or the cardiac cavity of recently dead NHPs. Samples from howler monkeys were centrifuged (2,000 g for 10 min) for obtaining plasma or serum samples stored at low temperature (N₂ or – 80 °C) until RNA extraction. Due to the small amount of blood in the cardiac cavity of dead marmosets, RNA extraction was performed from total blood frozen in dry ice immediately after collection in the field.

Human samples

Blood samples were obtained from patients for diagnosis procedures made at their respective municipal public health ambulatories. Serum samples of suspicious clinical YF cases were sent to the State Central Laboratory in Rio de Janeiro (LACEN-RJ) for molecular and serological analyses. Aliquots of serum samples of YFV laboratory-confirmed cases and negative for Zika virus, Chikungunya virus and Dengue virus infection were selected by the LACEN-RJ staff and stored at - 80 °C until RNA extraction.

YFV RNA extraction, screening for YFV infection by RT-PCR and Nucleotide Sequencing

YFV RNA from mosquito homogenates, blood, and serum samples were obtained by using the QIAamp Viral RNA kit (Qiagen). The RNA samples were eluted in 60 µL of AVE buffer and stored at - 80 °C until use. The YFV RNA was reverse transcribed using the Superscript IV First-Strand Synthesis System (Invitrogen) with random hexamers or specific primers at 25 °C for 10 min, 55 °C for 10 min and 80 °C for 10 min. Detection of YFV genome was performed by conventional PCR as described

elsewhere (15). Complete YFV genome amplification was carried out by conventional PCR using KAPA HiFi PCR kit (KAPABIOSYSTEMS) under the following conditions: 95 °C for 3 min, followed by 35 cycles at 98°C for 20 sec, 65°C for 15 sec and 72°C for 45 sec, and a final extension at 72 °C for 1 min 30 sec. The set of primers utilized in PCR and sequencing procedures are listed in the Supplementary files 2 and 3, respectively. Complete genome sequences were deposited in the GenBank database (Table 1). Amplified products were sequenced as previously described (15). The sequenced amplicons were analyzed, and contigs were assembled by using SeqMan Pro v8.1.5 (3), 414 (DNASTAR, Madison, WI). The sequences were manually edited and compared with other sequences available in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). The Molecular Evolutionary Genetics Analysis (MEGA) 7.0 program was used to calculate nucleotide and amino acid distances, as well as to explore the amino acid signatures observed in the YFV strains from the ongoing outbreak in Southeast Brazil.

Comparative modeling, optimization and MD simulation of NS3 pro and NS5 proteins

We performed the modeling of the NS2B-NS3 protein complex and NS5 protein of the 2017 outbreak YFV prototype (Genbank, KY885001) and the 2010 Venezuelan 10A strain (Genbank, KM388816). Initially, we used BLASTP program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>), defining the Protein Data Bank (PDB) as a search set, to select the template structures for comparative modeling. Three PDB structures were used as templates for the NS2B-NS3 protein complex. Template 2VBC (51% identity and 96% of coverage) corresponds to the crystal structure of the NS3 protease-helicase from dengue virus(38). Template 1YKS (96% identity and 70% of coverage) comprises the NS3 helicase domain of the YFV(39). Template 5GJ4 comprises the NS3 protease domain (56% identity and 27 % of coverage) and NS2B

peptide cofactor (42% identity and 31 % of coverage) of Zika virus(40). For the NS5 model generation, a single PDB structure from Japanese Encephalitis virus (4K6M) (19) was used as a template, which shares 60% identity with the YFV sequence. Template and target sequences were then aligned using the PSI-Coffee mode of T-Coffee program(41). One hundred homology models were generated using the standard “auto model” routine of Modeller version 9.18 (42) for each target sequence. Each model was optimized using the variable target function method (VTFM) optimization until accomplishing 300 iterations. Molecular dynamics (MD) optimization, in the slow level mode, was carried out. The full cycle was repeated two times to produce an optimized conformation of the model. The resulting modeled structures were selected according to their discrete optimized protein energy (DOPE) score. The GROMACS version 5.1.2 package(43), was used to carry out minimization using the AMBER99SB ILDN force field(44). A short minimization procedure of 150 steps (100 steps of steepest descent + 50 steps of conjugate-gradient) was performed. Initial and optimized models were evaluated by DOPE, Ramachandran plot and QMean server (Supplementary File 4) (45). The electrostatic potential analysis was conducted using the APBS program(46). Atomic partial charges and atomic radii parameters from the Amber force field were assigned using the PDB2PQR server (47). Figures of sequence alignments were rendered using ALINE (48), and 3D structures were generated using UCSF Chimera and PyMol.

Molecular dynamics simulations were carried out using the GROMACS package, with the AMBER99SB-ILDN force field. Protonation states were assigned using pdb2pqr software, and the zinc-coordinating cysteine residues were manually deprotonated. The models were further optimized prior to the MD runs, through 5.0×10^6 steps of the Steepest Descent (with and without heavy atom restraints) and Conjugate Gradients algorithms. The systems were then run, under an NPT ensemble, for 500 ps with restraints

and 2.0×10^5 ps without restraints. The V-rescale thermostat and Berendsen barostat were used for temperature and pressure control, respectively. The 2010 and 2017 strains and replicas were simulated at 297 and 310 K for a total of 8 runs and 1.6×10^6 ps. Analysis was made over the final 150 ns of the production runs.

PrM/E phylogenetic analysis

A 666-nucleotide sequence consisting of the last 108 nucleotides of the prM gene, the entire M gene (225 nucleotides), and the first 333 nucleotides of the E gene was used to established genotype the YFV strains, as previously described(5). In addition to the sequences obtained in the current study, a set of sequences of the prM/E junction fragment were selected using the Blast tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and all sequences were aligned using Molecular Evolutionary Genetics Analysis (MEGA) 7.0(49). The phylogenetic tree was generated by the Neighbor-joining method (50) using a matrix of genetic distances established under the Kimura-two parameter model (51). The robustness of each node was assessed by bootstrap resampling (2,000 replicates) (52). The homologous region (prM/E) of a dengue virus strain available at the GenBank database (AF349753) was used as an outgroup.

Selection pressure analyses

Two datasets of sequences were determined: a) Dataset South America I (Set SAI) with all complete CDS of South American YFV stains available in the GenBank plus sequences obtained in the current study (N=32); b) Dataset West Africa (Set WA), with complete CDS from West Africa genotype YFV strains available in the Genbank. The reason to generate these two datasets was that West African genotype has a closest genetic relationship with South America genotype I (5), and only two CDS from YFV strains belonging to South America genotype II are available in the GenBank to date. Datasets were aligned using Bioedit v7.2.3 (53) and were tested for positive selection by using the

online adaptive evolution server DATAMONKEY(54, 55). The ratio of non-synonymous to synonymous substitutions (dN/dS) per codon sites were estimated using four different methods, SLAC - Single Likelihood Ancestor Counting, FEL - Fixed Effects Likelihood(56), MEME - Mixed Effects Model of Evolution(57) and FUBAR - Fast Unbiased Bayesian Approximate(58). The analysis was run based on neighbor-joining tree and significant P-value (< 0.1). The automated model selection tool available in the server was used for selection of appropriate nucleotide substitution bias model for both datasets.

Evolutionary analyses

Complete coding regions sequences (CDS - 10,239 nt in length) of YFV of American origin (South America genotypes I and II), with a known date of isolation, were obtained from the GenBank database (www.ncbi.nlm.nih.gov). Retrieved sequences were aligned together with sequences obtained in the current study using MEGA 7.0 (49). The nucleotide substitution rate and time of the most recent common ancestor (T_{MRCA}) of American YFV strains were estimated using the Markov chain Monte Carlo (MCMC) algorithm implemented in the BEAST v1.8.3 package(59, 60) with BEAGLE (61) to improve run-time. The best-fit nucleotide substitution model, a relaxed uncorrelated lognormal molecular clock model(62), and a Bayesian Skyline coalescent tree prior(63) were used. The uncertainty of parameter estimates was assessed after excluding the initial 10% of the run by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively, using TRACER v1.6 program(64). TreeAnnotator v1.7.5 (60) and FigTree v1.4.0 (65) were used to summarize the posterior tree distribution and to visualize the annotated Maximum Clade Credibility (MCC) tree, respectively.

Acknowledgments

To Monique Albuquerque Motta for the help on mosquito identification; Iule de Souza Bonelli, Marcelo Quinteka Gomes, Lidiane S. R. Menezes, Nathália D. Furtado, Stephanie O. D. da Cruz and Adalgiza as Silva Rocha for technical assistance; Heloisa Diniz and Leônidas Leite Santos for the help with illustrations; Alexandre Otávio Chieppe, Ana Paula Martins Brandão, Carlos Augusto Ferandes (Secretaria Estadual de Saúde do Rio de Janeiro) Alessandro Pecego M. Romano (Grupo Técnico de Vigilância de Arboviroses) and Roberta Gomes de Carvalho (Programa Nacional de Controle da Dengue) Brazilian Ministry of Health, Gilsa Aparecida P. Rodrigues (Secretaria de Saúde do Estado do Espírito Santo), Gilton Luiz Almada (Centro de Informação Estratégica de Vigilância em Saúde-ES) for the access to epidemiological data and support for the field work; Marilza L. Lange (Secretaria Municipal de Saúde, Municipality of Domingos Martins), Luciano L. Salles (Vigilância em Saúde, Municipality of Ibatiba) Agenor Barbosa de Oliveira, Marcelo dos Santos Rosário, Mauro Cesar Louzada, Max Mauro dos Santos Coelho for their assistance in field collections in Espírito Santo. Alexandre Bezerra de Souza and Vicente Klonowski (Atalaia Park), Romenique de Lemos Araújo, Luiz Ribeiro Nogueira and Raquel Giri Faria (Environmental Guard) and Macaé prefecture for their support in collecting samples. Raul Henrique Rafael and Ana Luiza Quijada for access to the Guapimirim sample; Cassia Schitino de Carvalho Gomes and Paulo Vitor Araújo Gomes for access to the Carmo samples.

Competing interests

The authors declare that no competing interests exist.

REFERENCES

1. Barnett, E. D., Yellow fever: epidemiology and prevention. *Clin Infect Dis* **44**, 850-856 (2007), 10.1086/511869
2. Monath, T. P., Vasconcelos, P. F., Yellow fever. *J Clin Virol* **64**, 160-173 (2015), 10.1016/j.jcv.2014.08.030
3. Weaver, S. C., Reisen, W. K., Present and future arboviral threats. *Antiviral Res* **85**, 328-345 (2010), 10.1016/j.antiviral.2009.10.008
4. Vasconcelos, P. F., Monath, T. P., Yellow Fever Remains a Potential Threat to Public Health. *Vector Borne Zoonotic Dis* **16**, 566-567 (2016), 10.1089/vbz.2016.2031
5. Bryant, J. E., Holmes, E. C., Barrett, A. D., Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog* **3**, e75 (2007), 10.1371/journal.ppat.0030075
6. Mutebi, J. P., Wang, H., Li, L., Bryant, J. E., Barrett, A. D., Phylogenetic and evolutionary relationships among yellow fever virus isolates in Africa. *J Virol* **75**, 6999-7008 (2001), 10.1128/JVI.75.15.6999-7008.2001
7. Staples, J. E., Monath, T. P., Yellow fever: 100 years of discovery. *JAMA* **300**, 960-962 (2008), 10.1001/jama.300.8.960
8. Nunes, M. R., Palacios, G., Cardoso, J. F., Martins, L. C., Sousa, E. C., Jr., de Lima, C. P., Medeiros, D. B., Savji, N., Desai, A., Rodrigues, S. G., Carvalho, V. L., Lipkin, W. I., Vasconcelos, P. F., Genomic and phylogenetic characterization of Brazilian yellow fever virus strains. *J Virol* **86**, 13263-13271 (2012), 10.1128/JVI.00565-12

- 540 9. Vasconcelos, P. F., Bryant, J. E., da Rosa, T. P., Tesh, R. B., Rodrigues, S. G.,
541 Barrett, A. D., Genetic divergence and dispersal of yellow fever virus, Brazil.
542 *Emerg Infect Dis* **10**, 1578-1584 (2004), 10.3201/eid1009.040197
- 543 10. Romano, A. P., Costa, Z. G., Ramos, D. G., Andrade, M. A., Jayme Vde, S.,
544 Almeida, M. A., Vettorello, K. C., Mascheretti, M., Flannery, B., Yellow Fever
545 outbreaks in unvaccinated populations, Brazil, 2008-2009. *PLoS Negl Trop Dis* **8**,
546 e2740 (2014), 10.1371/journal.pntd.0002740
- 547 11. de Souza, R. P., Foster, P. G., Sallum, M. A., Coimbra, T. L., Maeda, A. Y.,
548 Silveira, V. R., Moreno, E. S., da Silva, F. G., Rocco, I. M., Ferreira, I. B., Suzuki,
549 A., Oshiro, F. M., Petrella, S. M., Pereira, L. E., Katz, G., Tengan, C. H., Siciliano,
550 M. M., Dos Santos, C. L., Detection of a new yellow fever virus lineage within
551 the South American genotype I in Brazil. *J Med Virol* **82**, 175-185 (2010),
552 10.1002/jmv.21606
- 553 12. Mir, D., Delatorre, E., Bonaldo, M., Lourenço-de-Oliveira, R., Vicente, A. C.,
554 Bello, G., Phylodynamics of Yellow Fever Virus in the Americas: new insights
555 into the origin of the 2017 Brazilian outbreak. *Scientific Reports* **7**, (2017),
556 10.1038/s41598-017-07873-7
- 557 13. Health, B. M. o., COES -Febre amarela. Informe 42/2-17, (2017),
558 [http://portalarquivos.saude.gov.br/images/pdf/2017/maio/26/COES-FEBRE-](http://portalarquivos.saude.gov.br/images/pdf/2017/maio/26/COES-FEBRE-AMARELA-INFORME-42.pdf)
559 [AMARELA-INFORME-42.pdf](http://portalarquivos.saude.gov.br/images/pdf/2017/maio/26/COES-FEBRE-AMARELA-INFORME-42.pdf)
- 560 14. Organization, P. A. H., Epidemiological Update -Yellow Fever, (2017),
561 [http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Item](http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=40841&lang=en)
562 [id=270&gid=40841&lang=en](http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=40841&lang=en)
- 563 15. Bonaldo, M. C., Gomez, M. M., Dos Santos, A. A., Abreu, F. V. S., Ferreira-de-
564 Brito, A., Miranda, R. M., Castro, M. G., Lourenco-de-Oliveira, R., Genome

analysis of yellow fever virus of the ongoing outbreak in Brazil reveals polymorphisms. *Mem Inst Oswaldo Cruz* **112**, 447-451 (2017), 10.1590/0074-02760170134

16. Luo, D., Vasudevan, S. G., Lescar, J., The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. *Antiviral Res* **118**, 148-158 (2015), 10.1016/j.antiviral.2015.03.014

17. Khromykh, A. A., Sedlak, P. L., Westaway, E. G., trans-Complementation analysis of the flavivirus Kunjin ns5 gene reveals an essential role for translation of its N-terminal half in RNA replication. *J Virol* **73**, 9247-9255 (1999), <https://www.ncbi.nlm.nih.gov/pubmed/10516033>)

18. Koonin, E. V., Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus. *J Gen Virol* **74** (Pt 4), 733-740 (1993), 10.1099/0022-1317-74-4-733

19. Lu, G., Gong, P., Crystal Structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog* **9**, e1003549 (2013), 10.1371/journal.ppat.1003549

20. El Sahili, A., Lescar, J., Dengue Virus Non-Structural Protein 5. *Viruses* **9**, (2017), 10.3390/v9040091

21. Laurent-Rolle, M., Morrison, J., Rajsbaum, R., Macleod, J. M. L., Pisanelli, G., Pham, A., Ayllon, J., Miorin, L., Martinez, C., tenOever, B. R., Garcia-Sastre, A., The interferon signaling antagonist function of yellow fever virus NS5 protein is activated by type I interferon. *Cell Host Microbe* **16**, 314-327 (2014), 10.1016/j.chom.2014.07.015

- 588 22. Cumberworth, S. L., Clark, J. J., Kohl, A., Donald, C. L., Inhibition of type I
589 interferon induction and signalling by mosquito-borne flaviviruses. *Cell*
590 *Microbiol* **19**, (2017), 10.1111/cmi.12737
- 591 23. Dorigatti, I., Hamlet, A., Aguas, R., Cattarino, L., Cori, A., Donnelly, C. A.,
592 Garske, T., Imai, N., Ferguson, N. M., International risk of yellow fever spread
593 from the ongoing outbreak in Brazil, December 2016 to May 2017. *Euro Surveill*
594 **22**, (2017), 10.2807/1560-7917.ES.2017.22.28.30572
- 595 24. Couto-Lima, D., Madec, Y., Bersot, M. I., Campos, S. S., Motta, M. A., Santos,
596 F. B. D., Vazeille, M., Vasconcelos, P., Lourenco-de-Oliveira, R., Failloux, A. B.,
597 Potential risk of re-emergence of urban transmission of Yellow Fever virus in
598 Brazil facilitated by competent Aedes populations. *Sci Rep* **7**, 4848 (2017),
599 10.1038/s41598-017-05186-3
- 600 25. Carpp, L. N., Galler, R., Bonaldo, M. C., Interaction between the yellow fever
601 virus nonstructural protein NS3 and the host protein Alix contributes to the release
602 of infectious particles. *Microbes Infect* **13**, 85-95 (2011),
603 10.1016/j.micinf.2010.10.010
- 604 26. Patkar, C. G., Kuhn, R. J., Yellow Fever virus NS3 plays an essential role in virus
605 assembly independent of its known enzymatic functions. *J Virol* **82**, 3342-3352
606 (2008), 10.1128/JVI.02447-07
- 607 27. Teramoto, T., Balasubramanian, A., Choi, K. H., Padmanabhan, R., Serotype-
608 specific interactions among functional domains of dengue virus 2 nonstructural
609 proteins (NS) 5 and NS3 are crucial for viral RNA replication. *J Biol Chem* **292**,
610 9465-9479 (2017), 10.1074/jbc.M117.775643
- 611 28. Lim, S. P., Koh, J. H., Seh, C. C., Liew, C. W., Davidson, A. D., Chua, L. S.,
612 Chandrasekaran, R., Cornvik, T. C., Shi, P. Y., Lescar, J., A crystal structure of

the dengue virus non-structural protein 5 (NS5) polymerase delineates interdomain amino acid residues that enhance its thermostability and de novo initiation activities. *J Biol Chem* **288**, 31105-31114 (2013), 10.1074/jbc.M113.508606

29. Zhao, Y., Soh, T. S., Zheng, J., Chan, K. W., Phoo, W. W., Lee, C. C., Tay, M. Y., Swaminathan, K., Cornvik, T. C., Lim, S. P., Shi, P. Y., Lescar, J., Vasudevan, S. G., Luo, D., A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog* **11**, e1004682 (2015), 10.1371/journal.ppat.1004682

30. Sironi, M., Forni, D., Clerici, M., Cagliani, R., Nonstructural Proteins Are Preferential Positive Selection Targets in Zika Virus and Related Flaviviruses. *PLoS Negl Trop Dis* **10**, e0004978 (2016), 10.1371/journal.pntd.0004978

31. Mathew, A., Townsley, E., Ennis, F. A., Elucidating the role of T cells in protection against and pathogenesis of dengue virus infections. *Future Microbiol* **9**, 411-425 (2014), 10.2217/fmb.13.171

32. Schlesinger, J. J., Foltzer, M., Chapman, S., The Fc portion of antibody to yellow fever virus NS1 is a determinant of protection against YF encephalitis in mice. *Virology* **192**, 132-141 (1993), 10.1006/viro.1993.1015

33. Wu, R. H., Tsai, M. H., Tsai, K. N., Tian, J. N., Wu, J. S., Wu, S. Y., Chern, J. H., Chen, C. H., Yueh, A., Mutagenesis of Dengue Virus Protein NS2A Revealed a Novel Domain Responsible for Virus-Induced Cytopathic Effect and Interactions between NS2A and NS2B Transmembrane Segments. *J Virol* **91**, (2017), 10.1128/JVI.01836-16

34. Le Breton, M., Meyniel-Schicklin, L., Deloire, A., Coutard, B., Canard, B., de Lamballerie, X., Andre, P., Rabourdin-Combe, C., Lotteau, V., Davoust, N.,

638 Flavivirus NS3 and NS5 proteins interaction network: a high-throughput yeast
639 two-hybrid screen. *BMC Microbiol* **11**, 234 (2011), 10.1186/1471-2180-11-234

640 35. Zou, J., Lee le, T., Wang, Q. Y., Xie, X., Lu, S., Yau, Y. H., Yuan, Z., Geifman
641 Shochat, S., Kang, C., Lescar, J., Shi, P. Y., Mapping the Interactions between the
642 NS4B and NS3 proteins of dengue virus. *J Virol* **89**, 3471-3483 (2015),
643 10.1128/JVI.03454-14

644 36. Faria, N. R., Azevedo, R., Kraemer, M. U. G., Souza, R., Cunha, M. S., Hill, S.
645 C., Theze, J., Bonsall, M. B., Bowden, T. A., Rissanen, I., Rocco, I. M., Nogueira,
646 J. S., Maeda, A. Y., Vasami, F., Macedo, F. L. L., Suzuki, A., Rodrigues, S. G.,
647 Cruz, A. C. R., Nunes, B. T., Medeiros, D. B. A., Rodrigues, D. S. G., Queiroz,
648 A. L. N., da Silva, E. V. P., Henriques, D. F., da Rosa, E. S. T., de Oliveira, C. S.,
649 Martins, L. C., Vasconcelos, H. B., Casseb, L. M. N., Simith, D. B., Messina, J.
650 P., Abade, L., Lourenco, J., Alcantara, L. C. J., de Lima, M. M., Giovanetti, M.,
651 Hay, S. I., de Oliveira, R. S., Lemos, P. D. S., de Oliveira, L. F., de Lima, C. P.
652 S., da Silva, S. P., de Vasconcelos, J. M., Franco, L., Cardoso, J. F., Vianez-Junior,
653 J., Mir, D., Bello, G., Delatorre, E., Khan, K., Creatore, M., Coelho, G. E., de
654 Oliveira, W. K., Tesh, R., Pybus, O. G., Nunes, M. R. T., Vasconcelos, P. F. C.,
655 Zika virus in the Americas: Early epidemiological and genetic findings. *Science*
656 **352**, 345-349 (2016), 10.1126/science.aaf5036

657 37. Consoli, R. A. G. B., Oliveira, R. L., *Principais mosquitos de importância*
658 *sanitária no Brasil*. (Editora FIOCRUZ, Rio de Janeiro, 1994), pp. 228.

659 38. Ferreira-de-Brito, A., Ribeiro, I. P., Miranda, R. M., Fernandes, R. S., Campos, S.
660 S., Silva, K. A., Castro, M. G., Bonaldo, M. C., Brasil, P., Lourenco-de-Oliveira,
661 R., First detection of natural infection of Aedes aegypti with Zika virus in Brazil

and throughout South America. *Mem Inst Oswaldo Cruz* **111**, 655-658 (2016),
10.1590/0074-02760160332

39. Luo, D., Xu, T., Hunke, C., Gruber, G., Vasudevan, S. G., Lescar, J., Crystal
structure of the NS3 protease-helicase from dengue virus. *J Virol* **82**, 173-183
(2008), 10.1128/JVI.01788-07

40. Wu, J., Bera, A. K., Kuhn, R. J., Smith, J. L., Structure of the Flavivirus helicase:
implications for catalytic activity, protein interactions, and proteolytic processing.
J Virol **79**, 10268-10277 (2005), 10.1128/JVI.79.16.10268-10277.2005

41. Phoo, W. W., Li, Y., Zhang, Z., Lee, M. Y., Loh, Y. R., Tan, Y. B., Ng, E. Y.,
Lescar, J., Kang, C., Luo, D., Structure of the NS2B-NS3 protease from Zika virus
after self-cleavage. *Nat Commun* **7**, 13410 (2016), 10.1038/ncomms13410

42. Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang,
J. M., Taly, J. F., Notredame, C., T-Coffee: a web server for the multiple sequence
alignment of protein and RNA sequences using structural information and
homology extension. *Nucleic Acids Res* **39**, W13-17 (2011), 10.1093/nar/gkr245

43. Webb, B., Sali, A., in *Protein Structure Prediction*, D. Kihara, Ed. (Springer New
York, New York, 2014), vol. 1137, pp. 1-15.

44. Abraham, M. K., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, H., Lindahl,
E., GROMACS: High performance molecular simulations through multi-level
parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015),
10.1016/j.softx.2015.06.001

45. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R.
O., Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB
protein force field. *Proteins* **78**, 1950-1958 (2010), 10.1002/prot.22711

686 46. Benkert, P., Kunzli, M., Schwede, T., QMEAN server for protein model quality
687 estimation. *Nucleic Acids Res* **37**, W510-514 (2009), 10.1093/nar/gkp322

688 47. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., McCammon, J. A., Electrostatics
689 of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad*
690 *Sci U S A* **98**, 10037-10041 (2001), 10.1073/pnas.181342398

691 48. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., Baker, N. A., PDB2PQR: an
692 automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.
693 *Nucleic Acids Res* **32**, W665-667 (2004), 10.1093/nar/gkh381

694 49. Bond, C. S., Schuttelkopf, A. W., ALINE: a WYSIWYG protein-sequence
695 alignment editor for publication-quality alignments. *Acta Crystallogr D Biol*
696 *Crystallogr* **65**, 510-512 (2009), 10.1107/S0907444909007835

697 50. Kumar, S., Stecher, G., Tamura, K., MEGA7: Molecular Evolutionary Genetics
698 Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874 (2016),
699 10.1093/molbev/msw054

700 51. Saitou, N., Nei, M., The neighbor-joining method: a new method for
701 reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425 (1987),
702 <https://www.ncbi.nlm.nih.gov/pubmed/3447015>)

703 52. Kimura, M., A simple method for estimating evolutionary rates of base
704 substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**,
705 111-120 (1980), <https://www.ncbi.nlm.nih.gov/pubmed/7463489>)

706 53. Felsenstein, J., Confidence Limits on Phylogenies: An Approach Using the
707 Bootstrap. *Evolution* **39**, 783-791 (1985), 10.1111/j.1558-5646.1985.tb00420.x

708 54. Hall, T. A., A User-Friendly Biological Sequence Alignment Editor and Analysis
709 Program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98
710 (1999),

711 55. Delpont, W., Poon, A. F., Frost, S. D., Kosakovsky Pond, S. L., Datamonkey 2010:
712 a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**,
713 2455-2457 (2010), 10.1093/bioinformatics/btq429

714 56. Pond, S. L., Frost, S. D., Datamonkey: rapid detection of selective pressure on
715 individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533 (2005),
716 10.1093/bioinformatics/bti320

717 57. Kosakovsky Pond, S. L., Frost, S. D., Not so different after all: a comparison of
718 methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-
719 1222 (2005), 10.1093/molbev/msi105

720 58. Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., Kosakovsky
721 Pond, S. L., Detecting individual sites subject to episodic diversifying selection.
722 *PLoS Genet* **8**, e1002764 (2012), 10.1371/journal.pgen.1002764

723 59. Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond,
724 S. L., Scheffler, K., FUBAR: a fast, unconstrained bayesian approximation for
725 inferring selection. *Mol Biol Evol* **30**, 1196-1205 (2013), 10.1093/molbev/mst030

726 60. Drummond, A. J., Rambaut, A., BEAST: Bayesian evolutionary analysis by
727 sampling trees. *BMC Evol Biol* **7**, 214 (2007), 10.1186/1471-2148-7-214

728 61. Drummond, A. J., Suchard, M. A., Xie, D., Rambaut, A., Bayesian phylogenetics
729 with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973 (2012),
730 10.1093/molbev/mss075

731 62. Suchard, M. A., Rambaut, A., Many-core algorithms for statistical phylogenetics.
732 *Bioinformatics* **25**, 1370-1376 (2009), 10.1093/bioinformatics/btp244

733 63. Drummond, A. J., Ho, S. Y., Phillips, M. J., Rambaut, A., Relaxed phylogenetics
734 and dating with confidence. *PLoS Biol* **4**, e88 (2006),
735 10.1371/journal.pbio.0040088

- 736 64. Drummond, A. J., Rambaut, A., Shapiro, B., Pybus, O. G., Bayesian coalescent
737 inference of past population dynamics from molecular sequences. *Mol Biol Evol*
738 **22**, 1185-1192 (2005), 10.1093/molbev/msi103
- 739 65. Rambaut, A., Suchard, M., Drummond, A., Tracer v1.6. , (2013),
740 <<http://tree.bio.ed.ac.uk/software/tracer/>>
- 741 66. Rambaut, A., FigTree v1.4., (2014), <<http://tree.bio.ed.ac.uk/software/figtree/>>

742

Figure Legends

Figure 1. The spatiotemporal spread of the yellow fever outbreak in Southeast Brazil from December 2016 to May 2017, and geographical origins of the yellow fever virus (YFV) samples according to states, river basins, hosts, and YFV sub-clades A (red) and B (green). The black square corresponds to the H189 sample whose only partial sequences of the YFV genome were obtained. Brazilian states: ES (Espírito Santo), MG (Minas Gerais), RJ (Rio de Janeiro) and SP: (São Paulo). Hatched areas correspond to the Great Metropolitan (GM) areas of Rio de Janeiro and Vitória.

Figure 2. Amino acid (aa) differences revealed by the alignment of the precursor polyproteins of 32 yellow fever (YF) viruses of the South America genotype I. On the left of the alignment data, the identification of lineage, sub-lineages and yellow fever virus (YFV) sequences are supplied. On the top of the alignment, the YF viral proteins positions are indicated along with the aa positions of amino acid differences. The orange-highlighted aa indicates the position of aa shared only by all YF sequences from the ongoing outbreak in Brazil. Amino acid residues highlighted in blue indicate aa changes present in YF strains from the current outbreak, first described in the current study. The “+” symbol indicates the sites under positive selection in the YF polyprotein.

Figure 3. Tridimensional models obtained by comparative modeling of the NS2B-NS3 protein complex. Cartoon and surface electrostatic potential representation of R121K substitution (A, B), whole complex (C, D), and E88D substitution (E, F). The molecular surface is colored according to electrostatic potential, where red, white and blue

correspond to acidic, neutral and basic potentials, respectively. NS2B is shown in green.

Thick black lines represent hydrogen bonds (A, B, E, F)

Figure 4. Tridimensional structural models obtained by comparative modeling of NS5 protein. (A) Cartoon representation of NS5 protein. Amino acid substitutions and binding site residues are shown in the sticks and colored according to the legend. (B) Cavities of NS5 protein. Amino acid substitution sites are shown in red.

Figure 5. Codons of the yellow fever virus under positive selection in South America genotype I (top), and West Africa genotype (bottom). The Y-axis represents normalized dN-dS (non-synonymous substitutions minus synonymous substitution), and the X-axis represents codon positions. The region between codon positions 2100 and 2850 is highlighted inside a gray rectangle. Positively selected sites are shown with the color code that appeared at the bottom of the figure.

Figure 6. Phylogenetic evolutionary analysis based on the yellow fever virus (YFV) complete coding region. (A) Time-scaled Bayesian MCC tree of YFV CDS. The color code is explained at the left of the figure. Names and accession numbers of the strains are shown in Fig. 6 – Fig. Supplement 1. (B) YFV from the ongoing Southeast Brazilian outbreak.

Figures –Figure Supplement legends

Figure 2- Figure Supplement 1. Phylogenetic analysis based on the prM/E junction region of yellow fever virus (YFV) strains analyzed in the current study and YFV sequences

retrieved from the National Centre for Biotechnology Information (NCBI). Only bootstrap values up to 70% are shown. YFV genotypes are shown at the right side of the figure. The scale bar at the bottom represents 0.1 substitutions per nucleotide position (nt.subst/site). YFV from the 2017 ongoing Southeast Brazilian outbreak are marked with a filled triangle (mosquito strains), filled square (human strains) and filled circle (non-human primates strains).

Figure 2- Figure Supplement 2. Phylogenetic analysis based on the NS3 (A) and NS5 (B) encoding region of yellow fever virus (YFV) strains analyzed in the current study and 21 YFV sequences retrieved from the National Centre for Biotechnology Information (NCBI). Only bootstrap values up to 70% are shown. The scale bar at the bottom represents 0.1 substitutions per nucleotide position (nt.subst/site). YFV from the 2017 ongoing Southeast Brazilian outbreak are marked with an empty circle. South America genotype I and sub-clade 1E are shown at the right side.

Figure 3 - Figure Supplement 1. Multiple sequence alignment of the NS3 protein sequence belonging to the 2017 Brazilian yellow fever virus (strain ES505), 2010 Venezuelan 10A strain and the three templates used in comparative modeling experiment. Black and gray filled positions of the alignment represent fully and partially conserved residues, respectively. Red outline highlights the amino acid found in the 2017 Brazilian strain. Green outline highlights the positions of the active site residues.

Figure 4 - Figure Supplement 1. Multiple sequence alignment of the NS5 protein sequence belonging to the 2017 Brazilian yellow fever virus (strain ES505), 2010 Venezuelan 10A strain and the template used in comparative modeling experiment. Black and gray filled positions of the alignment represent fully and partially conserved residues,

respectively. Red outline highlights the amino acid found in the 2017 Brazilian strain. Pink, yellow and green outlines highlight the positions of residues found in the active site, GTP binding site, and SAM binding site, respectively.

Figure 4 - Figure Supplement 2. The surface electrostatic potential of NS5 protein. The molecular surface is colored according to electrostatic potential, where red, white and blue correspond to acidic, neutral and basic potentials, respectively.

Figure 4 - Figure Supplement 3. Root mean square fluctuation plots of the final 150 ns of each NS5 MD production run. The systems were run in replicates with distinct random seeds for initial velocities generation. Temperatures of 297 and 310 K were used. The last column (Relative) represents the subtraction of the calculated fluctuations of 2010 from the 2017 values. The vertical dashed line marks the hinge domain. Fluctuations were calculated with GROMACS software and plotted using R.

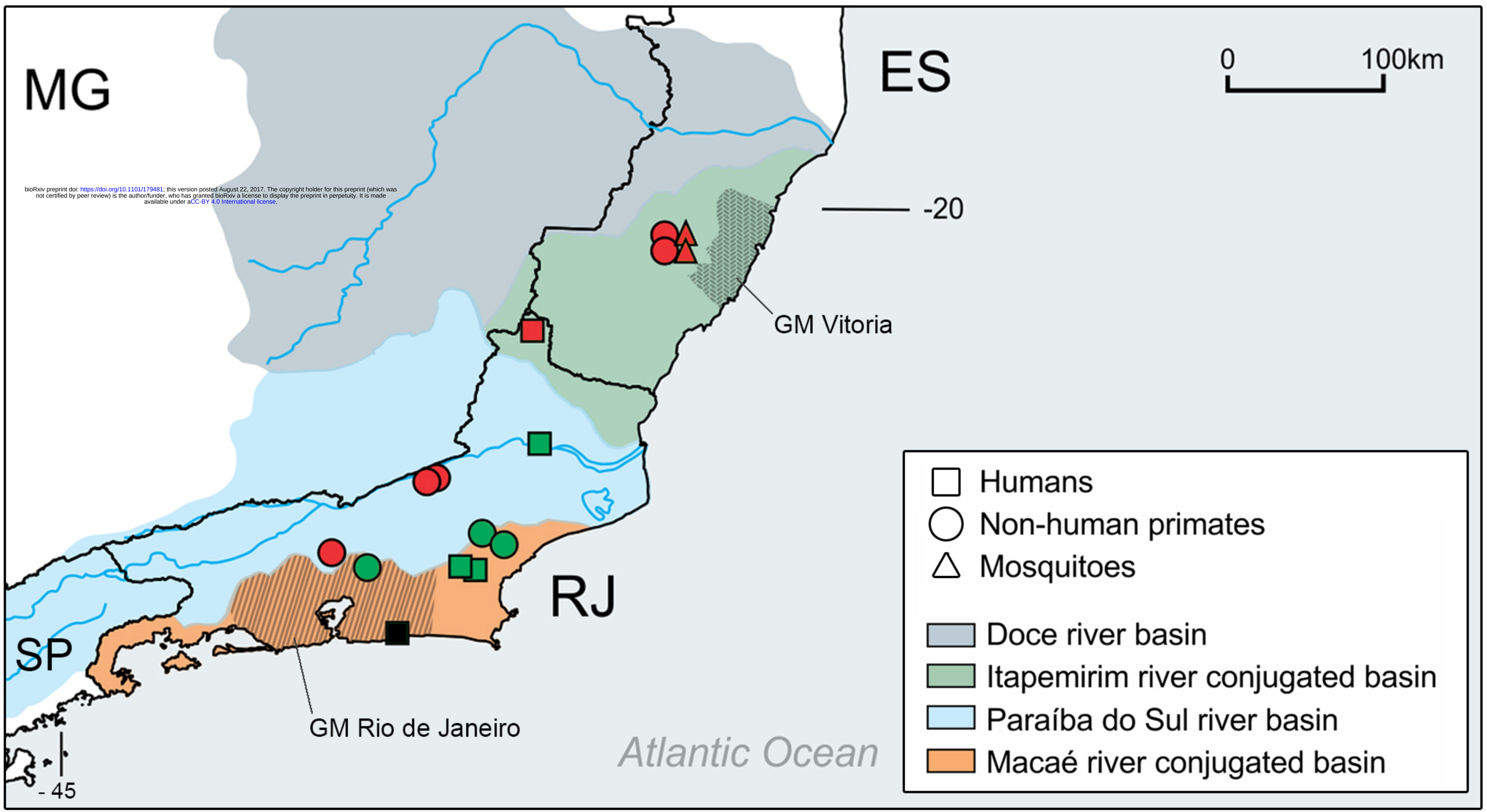
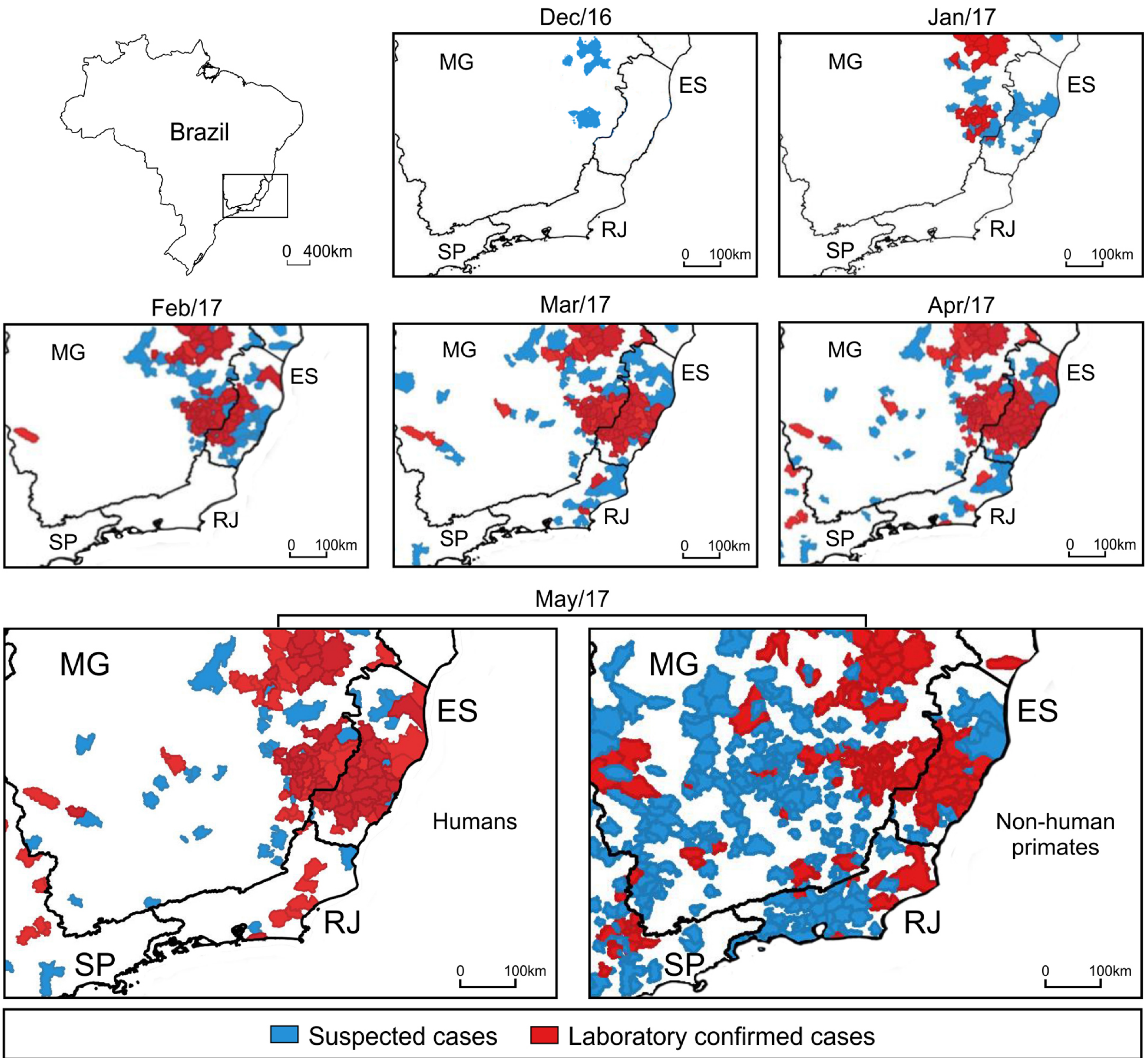
Figure 5 - Figure Supplement 1. Positively selected sites in yellow fever virus (YFV) polyprotein from South American genotype I (top) and West African (bottom). Sites, where the non-synonymous substitution occurred in YFV 2017 (top) are shown with a gray shadow. a- Sites supported by one method (FEL or MEME); b- sites supported by two methods (MEME/FUBAR or fel/fubar); c- site supported by three methods FEL/MEME/FUBAR.

Figure 6 – Figure supplement 1.. Yellow fever virus strains used in the Time-scaled Bayesian MCC tree.

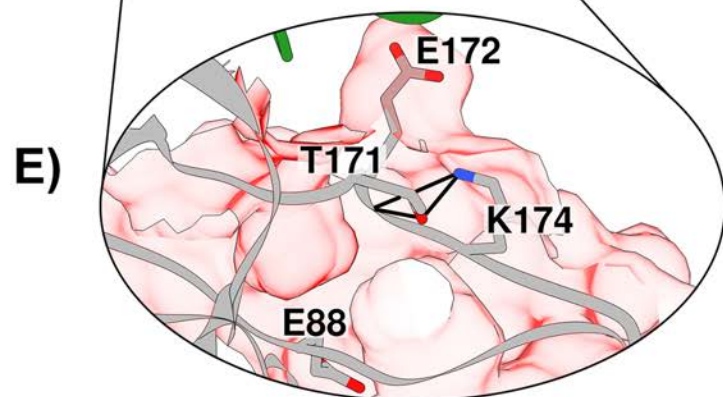
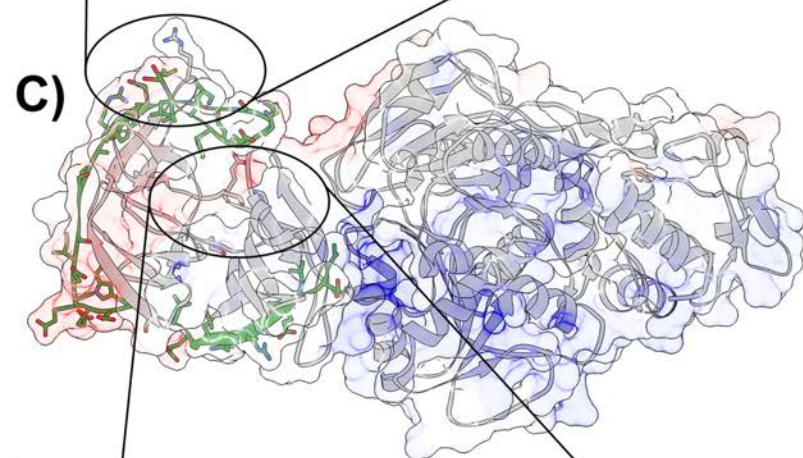
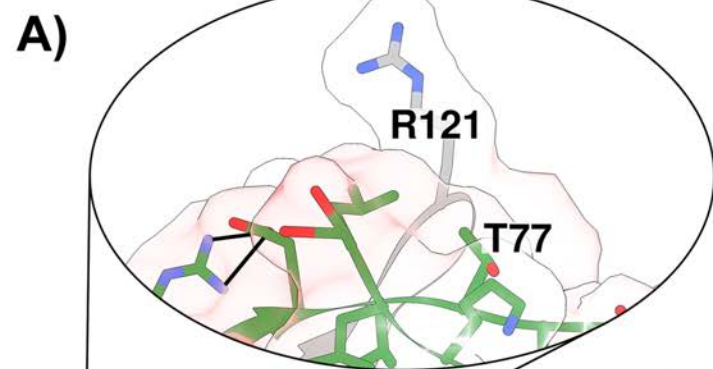
Table 1. YFV samples collected in the 2017 Brazilian outbreak

Host	Sample ID - (Genbank accession number)	Date of collection	Local	geographic coordinates
<i>Hg. leucocelaenus</i>	PA193 - (MF423373)	21/02/2017	Areinha, Domingos Martins - ES	20°17'08"S 40°50'15" W
<i>Hg. janthinomys</i>	PA196 - (MF423374)	23/02/2017	Areinha, Domingos Martins - ES	20°17'08"S 40°50'15"W
<i>Alouatta clamitans</i> (howler-monkey)	ES-504* - (KY885000)	20/02/2017	Areinha, Domingos Martins - ES	20°17'08"S 40°50'15"W
	ES-505* - (KY885001)	22/02/2017	Areinha, Domingos Martins - ES	20°17'08"S 40°50'15"W
	RJ87 - (MF423375)	04/04/2017	Atalaia, Macaé - RJ	22°18'31.6"S 42°00'01.7"W
	RJ94 - (MF423376)	13/04/2017	Cabeceira do Sana, Macaé - RJ	22°14'23.1"S 42°09'05.0"W
	RJ95 - (MF423377)	19/04/2017	Santa Fé, Carmo - RJ	21°53'05.0"S 42°32'29.7"W
	RJ96 - (MF423378)	19/04/2017	Santa Fé, Carmo - RJ	21°53'05.0"S 42°32'29.7"W
<i>Callithrix jacchus/penicillata</i> (marmoset)	RJ97- (MF538785)	21/04/2017	Araras, Petrópolis - RJ	22°23'51.1"S 43°10'56.5"W
	RJ104 - (MF538786)	05/06/2017	Caneca Fina, Guapimirim - RJ	22°29'35.9"S 42°56'58.9"W
Human cases	H189 **	18/04/2017	Bananal, Maricá - RJ	22°55'25.3"S 42°43'17.1"W
	H190 - (MF538782)	16/03/2017	São Fidélis - RJ	21°38'17.2"S 41°45'49.2"W
	H191 - (MF538783)	18/03/2017	Casimiro de Abreu - RJ	22°29'10.5"S 42°12'06.2"W
	H196 - (MF538784)	26/02/2017	Porciúncula - RJ	20°49'17.5"S 41°54'38.6"W
	H199 - (MF434851)	25/04/2017	Silva Jardim - RJ	22°27'42.9"S 42°18'28.6"W

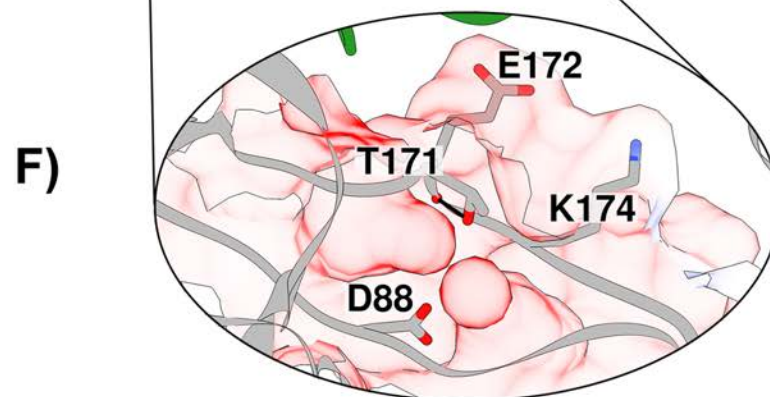
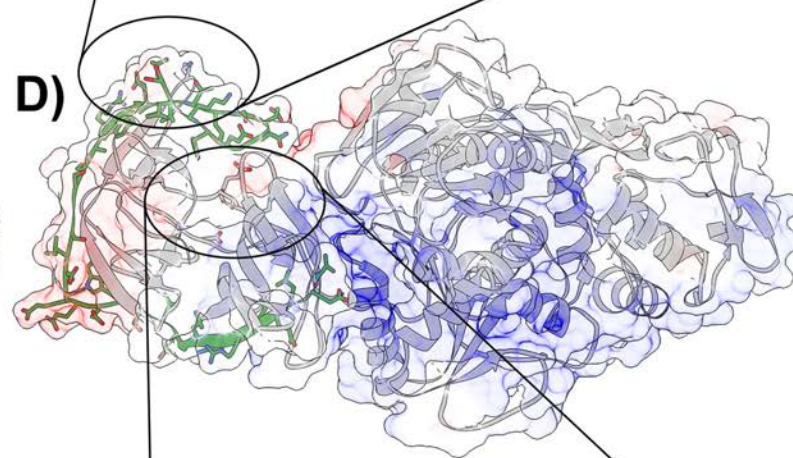
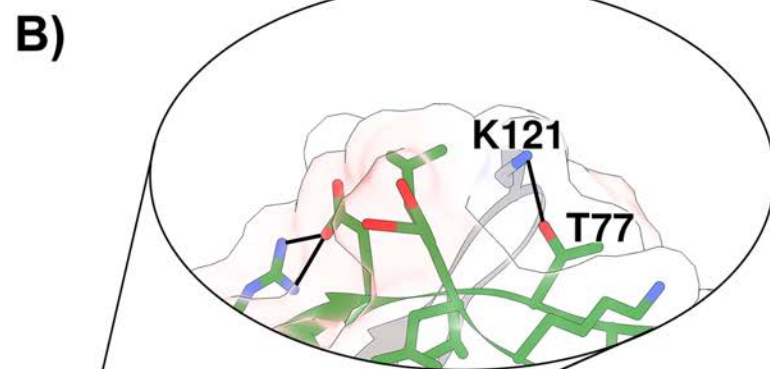
*YFV ES504 and ES505 previously described (13); ** partial genome sequence



YFV 2010



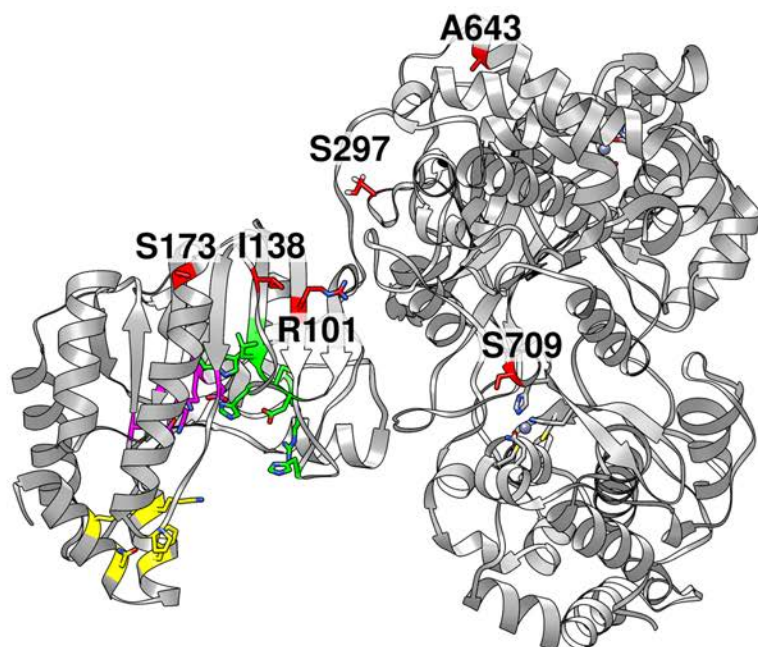
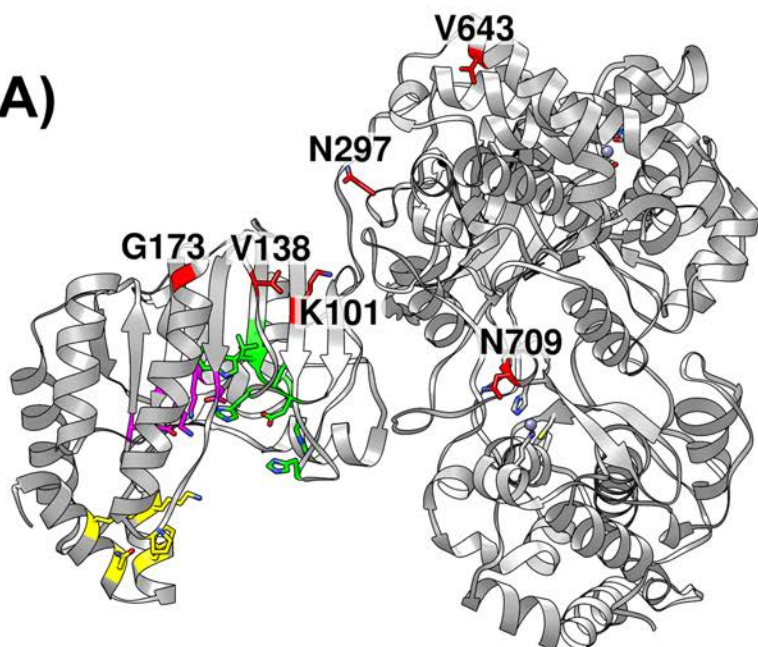
YFV 2017



YFV 2010

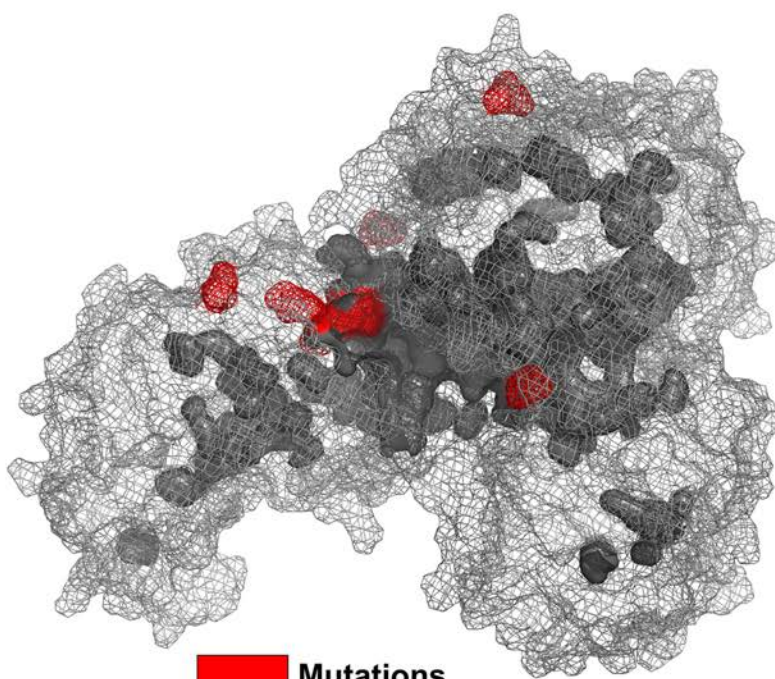
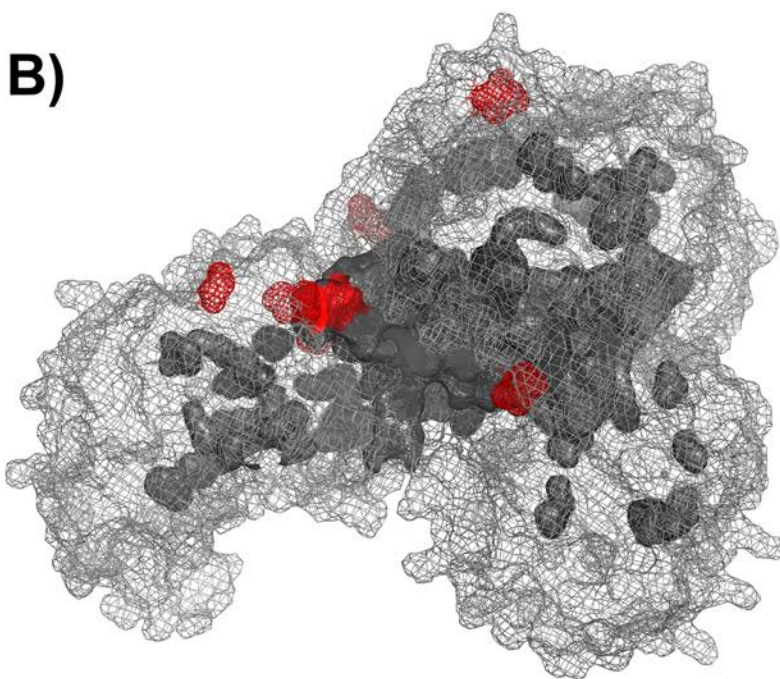
YFV 2017

A)



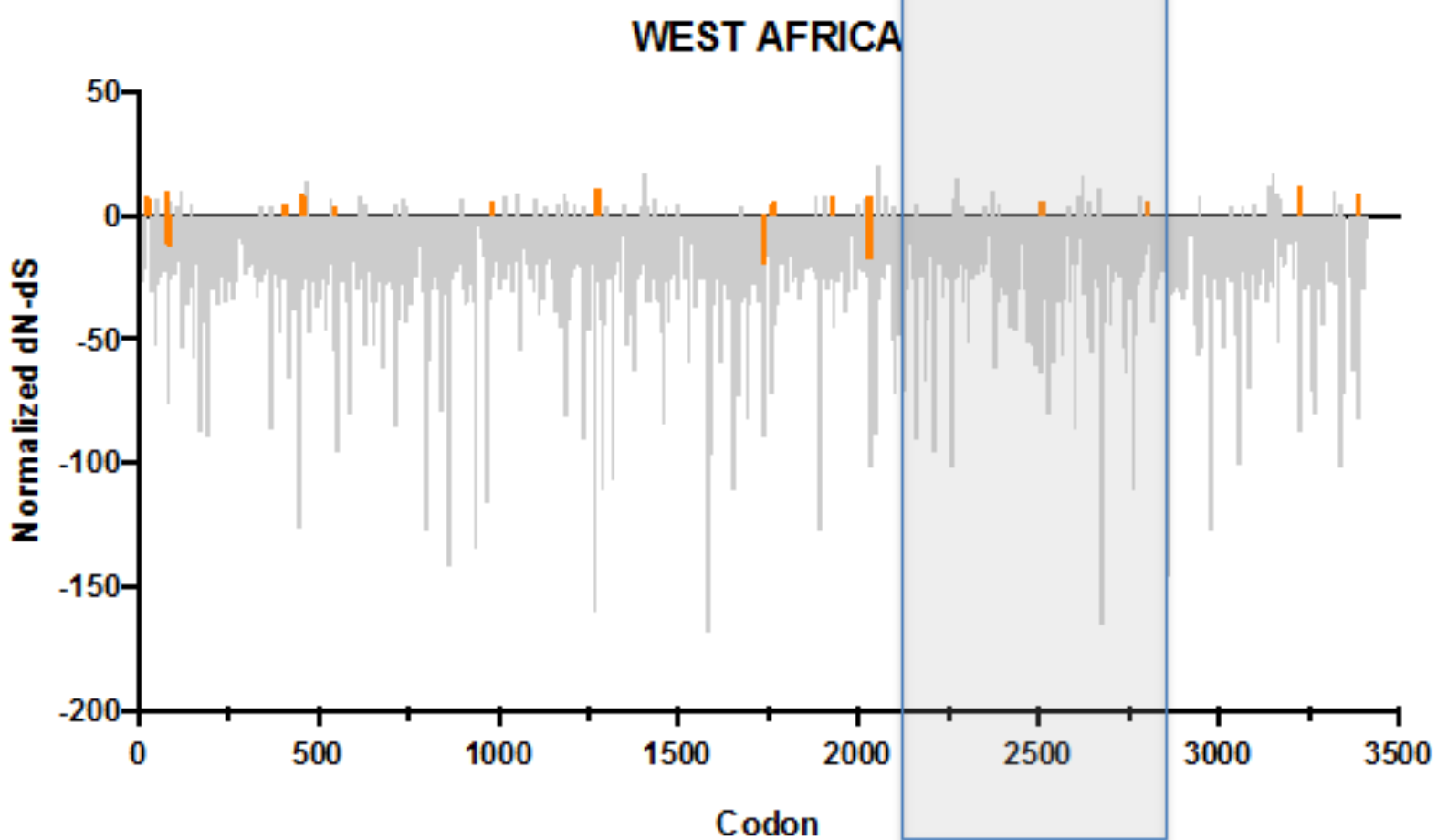
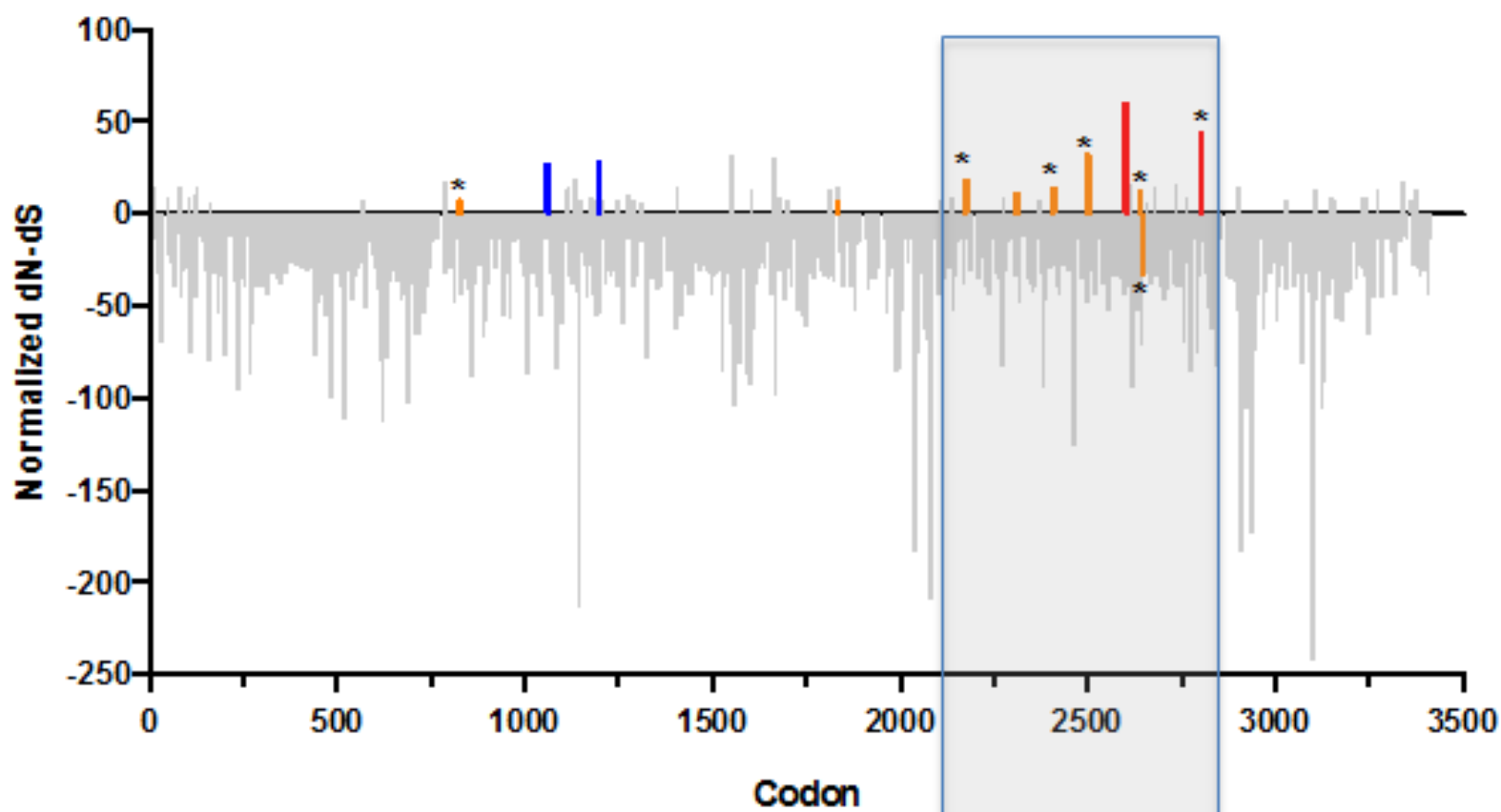
Red box: Mutations Magenta box: Active site Yellow box: GTP binding site Green box: SAM binding site

B)

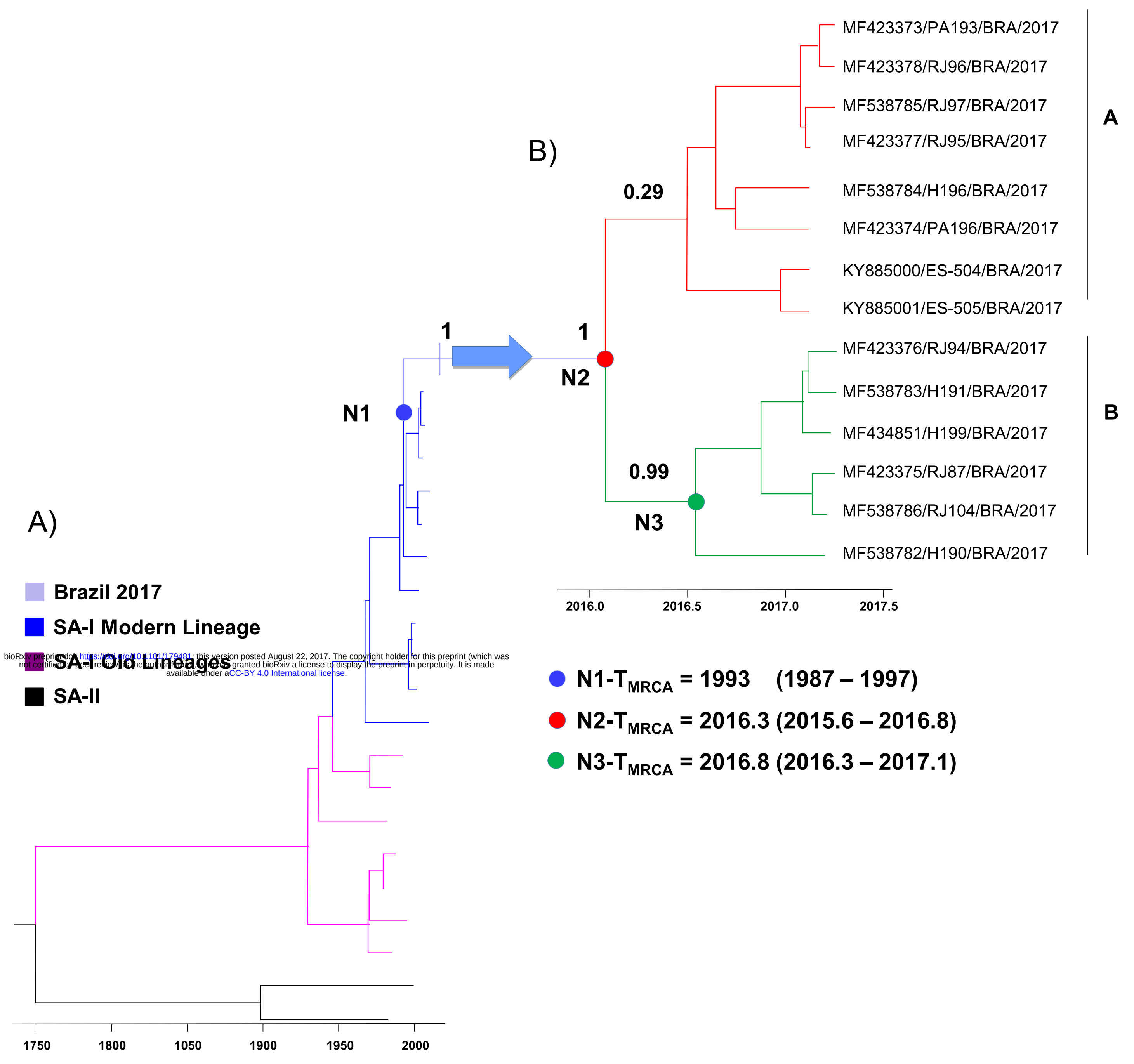


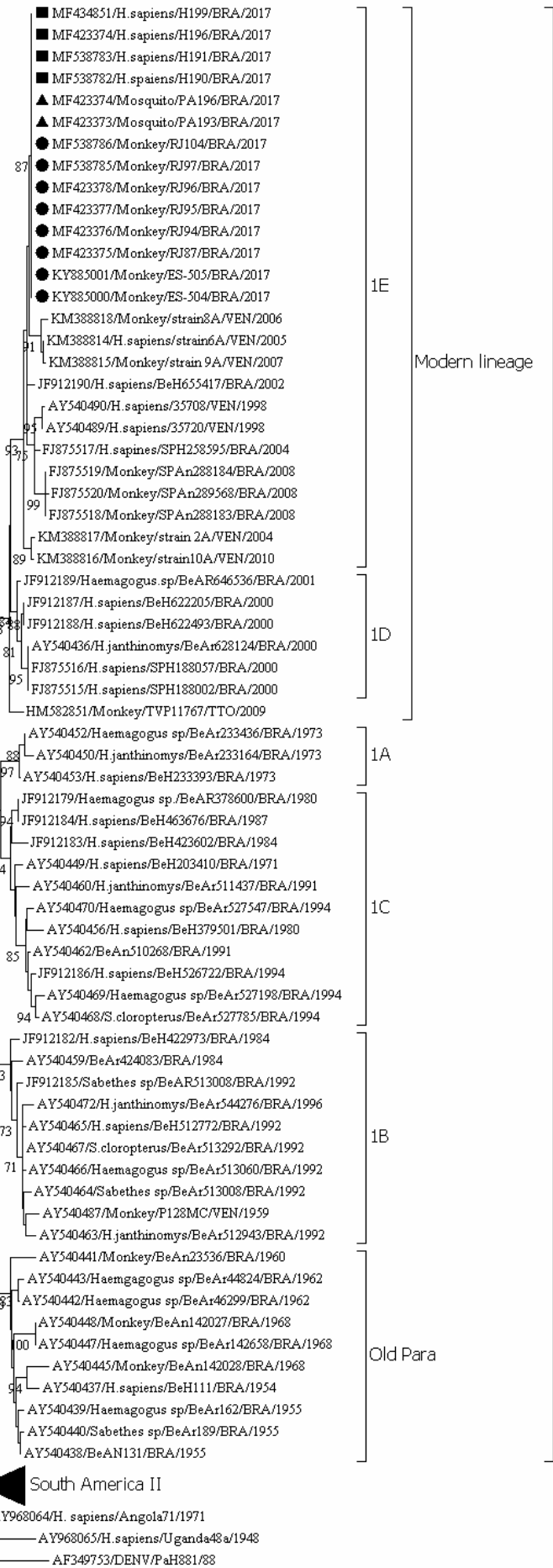
Red box: Mutations

SOUTH AMERICA-I

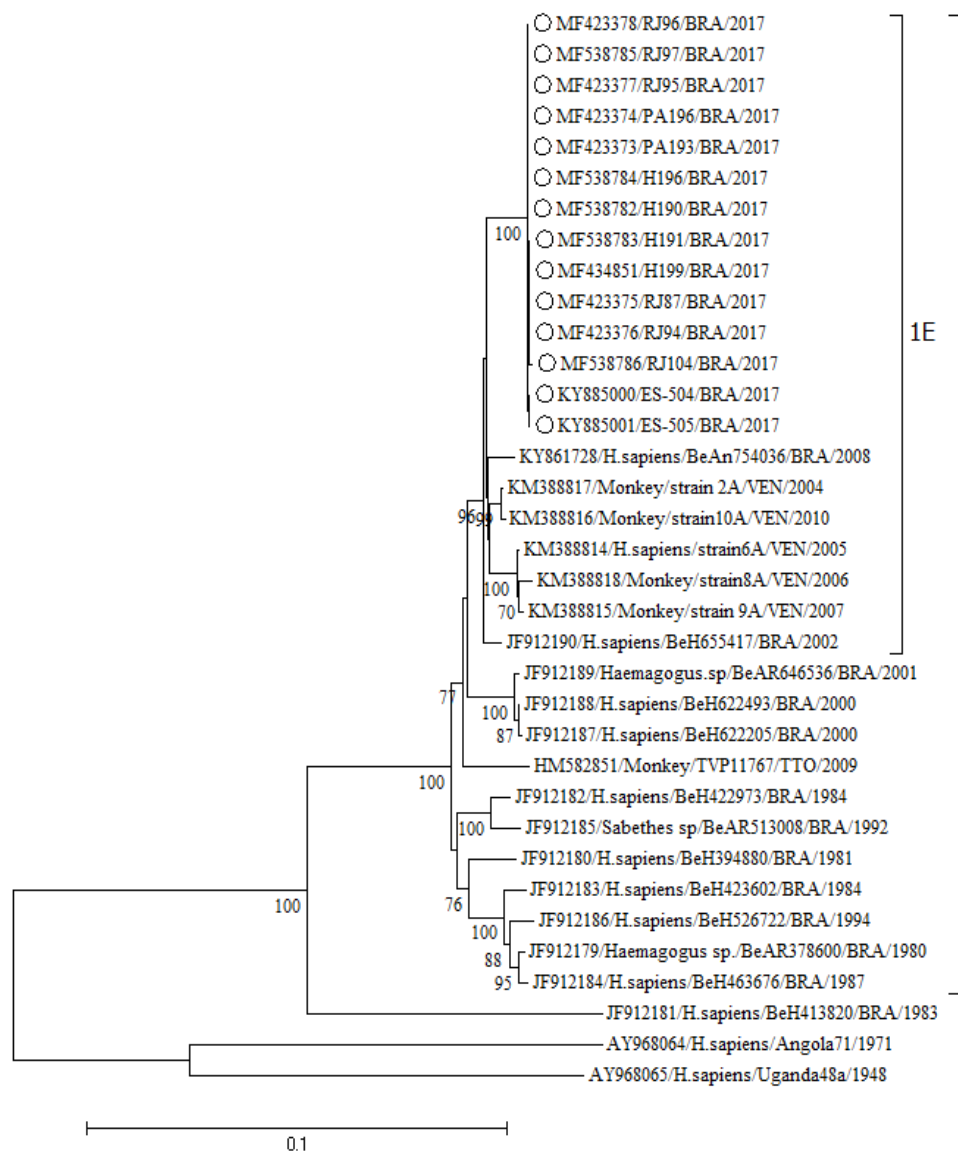


MEME FEL MEME/FEL/FUBAR

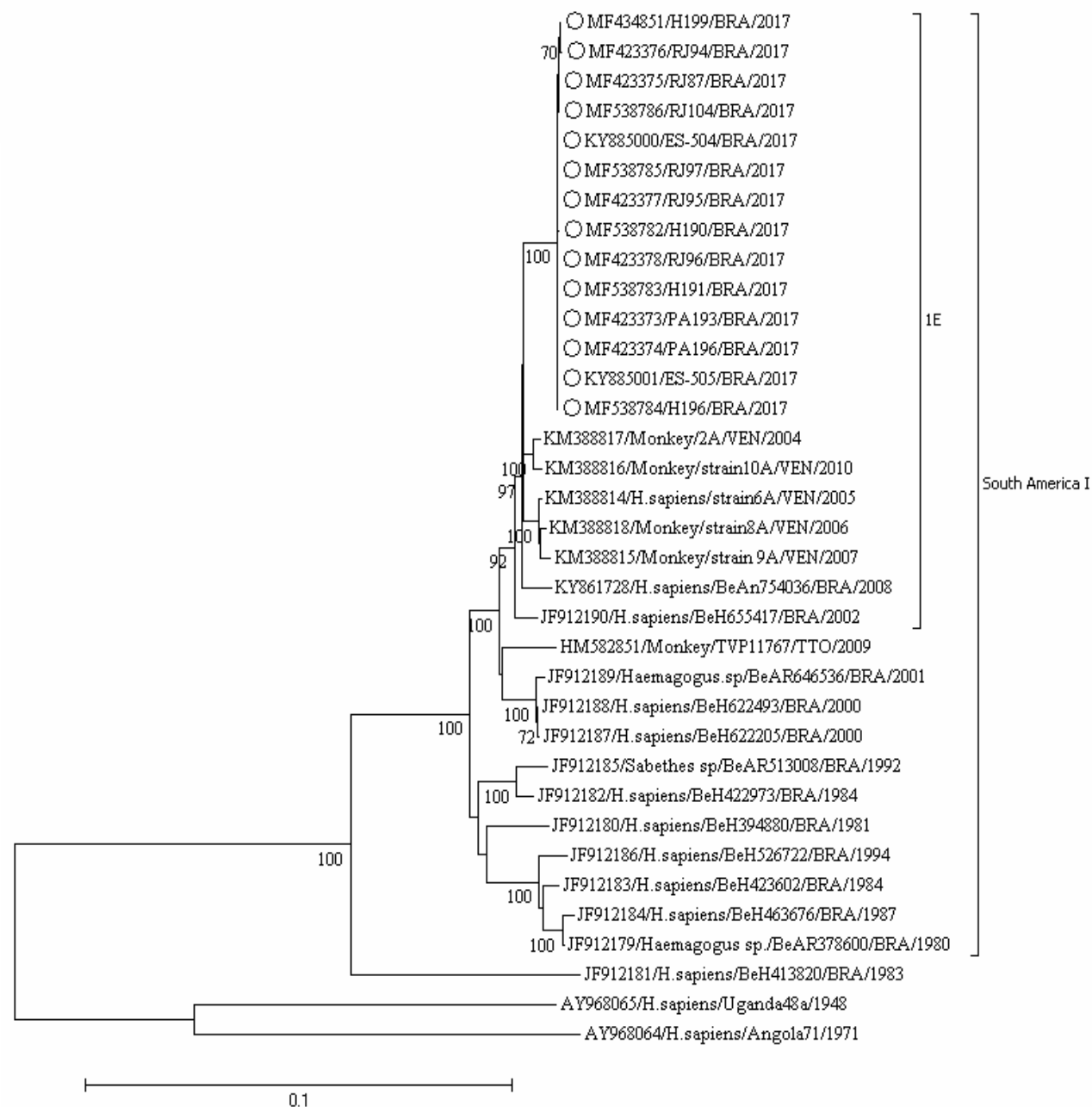




0.1



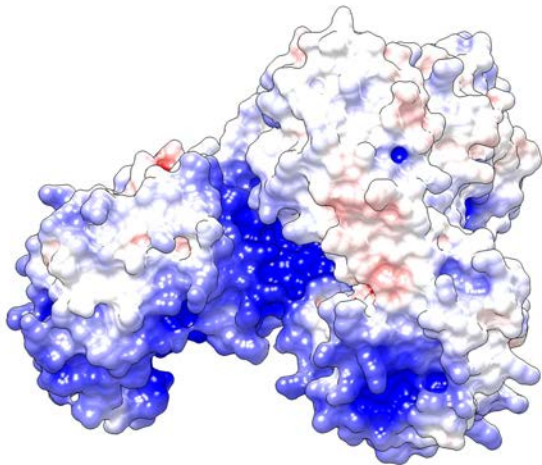
South America I



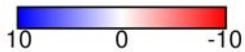
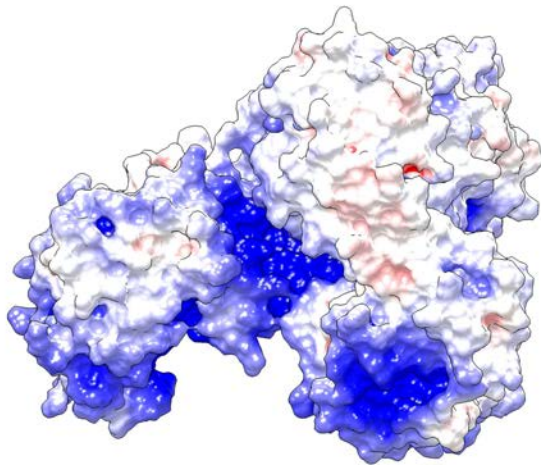
bioRxiv preprint doi: <https://doi.org/10.1101/179481>; this version posted August 22, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ES-505/BRA/2017	5691215182124273033363942454851545760	GKTLGEVWKRELNLLDKQQFELYKRTDIVEVDRDTARRHLAEGKVDTGVAVSRGTA	60
KM388816/Monkey/strain_10A/VEN/2010		GKTLGEVWKRELNLLDKQQFELYKRTDIVEVDRDTARRHLAEGKVDTGVAVSRGTA	60
4K6M_JEV		GRTLGEQWKEKLNAMSREEFFFKYRREAIIIEVDRTEARRARRENNIVGGHPVSRGSA	60
ES-505/BRA/2017	63666972757881848790939699102105108111114	KLRWFHERGYVKLEGRVTDLGCGRGGWCYAAAQREVSGVVRGFTTLGKEGHEKPMNV	116
KM388816/Monkey/strain_10A/VEN/2010		KLRWFHERGYVKLEGRVTDLGCGRGGWCYAAAQREVSGVVRGFTTLGKEGHEKPMNV	116
4K6M_JEV		KLRWLVEKGFVSPIGKVIDLGCGRGGWSYAAATLKKVQEVVRGYTKGGAGHEEPMLM	116
ES-505/BRA/2017	117120123126129132135138141144147150153156159162165168171	QSLGWNIIITFKDKTDLVHRLPEIKCDTLLCDIGESSPSSVTEGERTMRVLDTVEKWL	172
KM388816/Monkey/strain_10A/VEN/2010		QSLGWNIIITFKDKTDLVHRLPEIKCDTLLCDIGESSPSSVTEGERTMRVLDTVEKWL	172
4K6M_JEV		QSYGWNLVSLKSGVDVIFYKPSEPSDTLFCDIGESSPSPEVEEQRTLRLVLEMTSDWL	172
ES-505/BRA/2017	174177180183186189192195198201204207210213216219222225228	SCGVESFCVKVLAPYMPDVLEKLELLQRRFGGTVIRNPLSRNSTHEMYYVSGARSN	228
KM388816/Monkey/strain_10A/VEN/2010		SCGVESFCVKVLAPYMPDVLEKLELLQRRFGGTVIRNPLSRNSTHEMYYVSGARSN	228
4K6M_JEV		HRGPREFCIKVLCPYMPKVIEKMEVLQRRFGGGLVRLPLSRNSNHEMYWVSGAAGN	228
ES-505/BRA/2017	231234237240243246249252255258261264267270273276279282	ITFTVNQTSRLLMRRMRRTPTGK.VTLEADVILPIGTRSVETDKGPLDRAAIEERVE	283
KM388816/Monkey/strain_10A/VEN/2010		ITFTVNQTSRLLMRRMRRTPTGK.VTLEADVILPIGTRSVETDKGPLDRAAIEERVE	283
4K6M_JEV		VHAVNNTSGVLLGRMDRTVWRGPKYEEEDVNLGSGTRA V GK...GSNQEKIKKRIQ	281
ES-505/BRA/2017	285288291294297300303306309312315318321324327330333336339	RIKSEYTATWFHDSNDNPNYRTWHYCGSYVTRTSGSAASMINGVIKILTYPWDRIEEV	339
KM388816/Monkey/strain_10A/VEN/2010		RIKSEYTATWFHDSNDNPNYRTWHYCGSYVTRTSGSAASMINGVIKILTYPWDRIEEV	339
4K6M_JEV		KLKEEFATTWHKDPHPYRTWTYHGSYEVKATGSASSLVNGVVKLMSKPWDAIANV	337
ES-505/BRA/2017	342345348351354357360363366369372375378381384387390393	TRMAMTDTTPFQQQRVFKEKVDTRAKDPPAGTRKIMKVVNRLFRHLAREKNPRLC	395
KM388816/Monkey/strain_10A/VEN/2010		TRMAMTDTTPFQQQRVFKEKVDTRAKDPPAGTRKIMKVVNRLFRHLAREKNPRLC	395
4K6M_JEV		TTMAMTDTTPFQQQRVFKEKVDTKAPEPPAGAKEVLNETTNWLWAYLSREKRPRLC	393
ES-505/BRA/2017	396399402405408411414417420423426429432435438441444447450	TKEEFI AKVRSHAAI GAFLEE QEQWKTANEAVQDPKFWELVDEERRLHQQGR CRT C	451
KM388816/Monkey/strain_10A/VEN/2010		TKEEFI AKVRSHAAI GAFLEE QEQWKTANEAVQDPKFWELVDEERRLHQQGR CRT C	451
4K6M_JEV		TKEEFIKKVNSNAALGAVFAEQNQWSTAREAVDDPRFWEMVDEERENHLRGECHTC	449
ES-505/BRA/2017	453456459462465468471474477480483486489492495498501504507	VYNMMGKREKKLSEFGKAKGSRAIWYMWLGARYLEFEALGFLNEDHWA SRENSGGG	507
KM388816/Monkey/strain_10A/VEN/2010		VYNMMGKREKKLSEFGKAKGSRAIWYMWLGARYLEFEALGFLNEDHWA SRENSGGG	507
4K6M_JEV		IYNMMGKREKKPGSEFGKAKGSRAIWF MWLGARYLEFEALGFLNEDHWL SRENSGGG	505
ES-505/BRA/2017	510513516519522525528531534537540543546549552555558561	VEGIGLQYLG YVIRDLAALEGGGFYADDTAGWDTRITEADLDDEQEILNYMSPHHR	563
KM388816/Monkey/strain_10A/VEN/2010		VEGIGLQYLG YVIRDLATLEGGGFYADDTAGWDTRITEADLDDEQEILNYMSPHHR	563
4K6M_JEV		VEGSGVQKLGYILRDLIAGKQGGKMYADDTAGWDTRITRTDLENEAKVLELLDGEHR	561
ES-505/BRA/2017	564567570573576579582585588591594597600603606609612615618	KLALAVMEMTYKNKVVKVLRPAPGGKAYMDVISRRDQRGSGQVVTYALNTITNLKV	619
KM388816/Monkey/strain_10A/VEN/2010		KLALAVMEMTYKNKVVKVLRPAPGGKAYMDVISRRDQRGSGQVVTYALNTITNLKV	619
4K6M_JEV		MLARAIIELTYRHKVVKVMRPA AEGKTVM DVVISRE DQRGSGQVVTYALNTFTNI AV	617
ES-505/BRA/2017	621624627630633636639642645648651654657660663666669672675	QLIRMAEAEMVIHHQH VQDCDDTALT KLEAWLAEHGC DRLKRMAVSGDDCVVRPID	675
KM388816/Monkey/strain_10A/VEN/2010		QLIRMAEAEMVIHHQH VQDCDDTALT KLEAWLAEHGC DRLKRMAVSGDDCVVRPID	675
4K6M_JEV		QLVRLMEAEGVIGPQHLEQLPRKNKIAVRTWL FENGEEERVTRMAISGDDCVVKPLD	673
ES-505/BRA/2017	678681684687690693696699702705708711714717720723726729	DRFGLALSHLNAMSKVRKDI SEWQPSKGWDDWE SV PFC SHHFHELQLKDGRRI VVP	731
KM388816/Monkey/strain_10A/VEN/2010		DRFGLALSHLNAMSKVRKDI SEWQPSKGWDDWE NV PFC SHHFHELQLKDGRRI VVP	731
4K6M_JEV		DRFATALHF LNAMSKVRKDIQEWKPSHGWHDWQQVPFC SNHFQEIVMKDGRSIVVP	729
ES-505/BRA/2017	732735738741744747750753756759762765768771774777780783786	CRDQDEL VGRGRVSPGN GWM I KETACLSKAYANMWSLMYFHKRDMRLLSLAVSSAV	787
KM388816/Monkey/strain_10A/VEN/2010		CRDQDEL VGRGRVSPGN GWM I KETACLSKAYANMWSLMYFHKRDMRLLSLAVSSAV	787
4K6M_JEV		CRGQDEL IGRARI SPGAGWNVKDTACLA KAYAQMWL LLYFHRRLRLMANA ICSAV	785
ES-505/BRA/2017	789792795798801804807810813816819822825828831834837840843	PTSWVPQGR TTWSVHGKGEWMTTEDML EVWNRVWI TNNPHMQDKTTVKEW RDIPYL	843
KM388816/Monkey/strain_10A/VEN/2010		PTSWVPQGR TTWSVHGKGEWMTTEDML EVWNRVWI TNNPHMQDKTTVKEW RDIPYL	843
4K6M_JEV		PVDWVPTGRTSWSIHSKGEWMTTEDMLQVWNRVWI EENEWMMDKTPITSWTDVPYV	841
ES-505/BRA/2017	846849852855858861864867870873876879882885888891894	TKRQDKLCGSLIGMTNRATWASHIHLV IHRIRTLIGKERYTDYLTVM DRY S	894
KM388816/Monkey/strain_10A/VEN/2010		TKRQDKLCGSLIGMTNRATWASHIHLV IHRIRTLIGKERYTDYLTVM DRY S	894
4K6M_JEV		GKREDIWCGSLIGTRS RATWAENIYAAINQVRAVIGKENYVDYMTSL RRYE	892

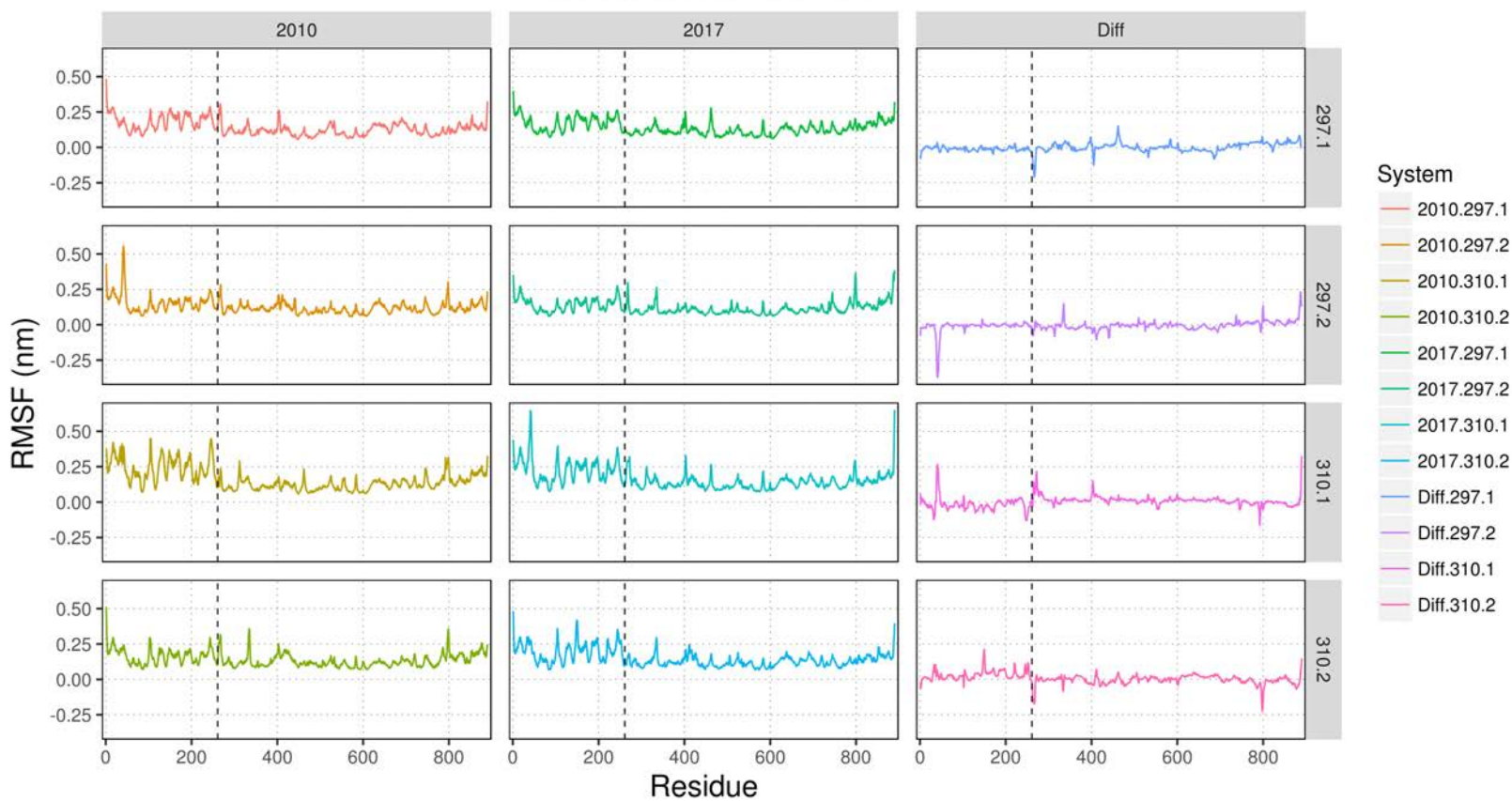
YFV 2010



YFV 2017



Backbone RMSF



Seq Id / Strain / Country / Year		NS1		NS2A	NS3	NS4A	NS4B				NS5					
		826 ^a	1062 ^a	1197 ^a	1835 ^a	2176 ^a	2311 ^a	2408 ^a	2502 ^a	2503 ^a	2601 ^c	2640 ^a	2647 ^a	2803 ^b		
South America genotype I	Old lineage	1B	JP912180/BeH394880/BRA/1981	A	N	N	T	I	H	N	K	T	R	R	D	N
			JP912182/BeH422973/BRA/1984	A	S	N	T	I	H	N	K	T	R	R	D	N
			JP912185/BeAR513008/BRA/1992	A	S	N	T	I	H	N	K	T	R	R	D	N
		1C	JP912179/BeAR378600/BRA/1980	A	S	N	T	I	H	N	N	T	K	R	D	N
			JP912183/BeH423602/BRA/1984	A	S	N	T	I	H	N	N	T	R	R	D	N
			JP912184/BeH463676/BRA/1987	A	S	N	T	I	H	N	N	T	K	R	D	N
		1D	JP912186/BeH526722/BRA/1994	A	S	S	T	I	H	N	K	T	K	R	D	N
			JP912188/BeH622493/BRA/2000	A	S	N	T	I	H	N	K	T	R	R	D	D
			JP912187/BeH622205/BRA/2000	A	S	N	T	I	H	N	K	T	R	R	D	D
	Modern lineage	1E	JP912189/BeAR646536/BRA/2001	A	S	N	T	I	H	N	K	T	R	P	V	D
			JP912190/BeH655417/BRA/2002	A	S	N	T	I	H	N	K	T	R	R	D	N
			KM388817/strain 2A/VEN/2004	A	S	N	T	I	H	N	K	T	R	R	D	N
			KM388814/strain6A/VEN/2005	A	S	N	T	I	H	N	K	I	R	R	D	N
			KM388818/strain8A/VEN/2006	A	S	N	K	I	L	S	N	T	R	R	D	S
			KM388815/strain 9A/VEN/2007	A	S	S	T	I	H	S	K	T	R	R	D	N
			KY861728/BeAn754036/BRA/2008	A	N	N	T	I	H	N	K	T	K	R	D	N
			KM388816/strain10A/VEN/2010	A	S	N	T	I	H	N	K	T	R	R	D	N
			KY885000/ES-504/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			KY885001/ES-505/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			MF423375/RJ87/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			MF423376/RJ94/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			MF423377/RJ95/BRA/2017	A	S	N	T	V	H	N	K	T	R	R	D	S
			MF423378/RJ96/BRA/2017	A	S	N	T	V	H	N	K	T	R	R	D	S
			MF538785/RJ97/BRA/2017	A	S	N	T	V	H	N	K	T	R	R	D	S
			MF538786/RJ104/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			MF423373/PA193/BRA/2017	A	S	N	T	V	H	N	K	T	R	R	D	S
			MF423374/PA196/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			MF538782/H190/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
			MF538783/H191/BRA/2017	A	S	N	T	V	H	N	K	T	R	R	D	S
			MF434851/H199/BRA/2017	S	S	N	T	I	H	N	K	T	R	R	D	S
			MF538784/H196/BRA/2017	A	S	N	T	I	H	N	K	T	R	R	D	S
	TT		HM582851/TVP11767/TTO/2009	A	S	N	T	I	H	N	K	T	K	R	D	N

West African YFV strains	NS1																			NS2A		NS3		NS4B		NS5			
	C			E				NS1	NS2A	NS3				NS5															
	25 ^d	79 ^d	86 ^d	407 ^d	409 ^d	456 ^d	547 ^d	983 ^d	1277 ^d	1738 ^d	1764 ^d	1928 ^d	2029 ^d	2030 ^d	2508 ^d	2801 ^d	3227 ^d	3388 ^d											
	AY640589/Asibi/GHA/1927	N	A	R	A	S	E	M	H	T	A	V	K	D	L	S	Y	Q	K										
	JX898869/DakArAmt7/CIV/1973	N	A	R	A	S	E	M	H	V	A	I	K	D	L	R	Y	Q	K										
	U54798/85-82H/CIV/1982	N	A	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	Q	K										
	JX898868/HD117294/SEN/1995	N	A	R	A	S	E	M	H	T	G	I	K	D	L	R	H	Q	K										
	JX898871/ArD114896/SEN/1995	N	A	R	A	S	E	M	H	T	A	I	K	D	L	T	Y	Q	K										
	JX898872/ArD114972/SEN/1995	N	A	R	A	S	E	M	H	T	A	I	K	D	L	T	Y	Q	K										
	JX898870/ArD121040/SEN/1996	S	V	K	A	S	E	M	N	T	A	I	K	D	L	R	Y	Q	K										
	AY603338/Ivory Coast/CIV/1999	N	A	R	A	S	E	M	H	T	A	I	K	D	Q	R	Y	H	K										
	JX898873/ArD149214/SEN/2000	N	A	R	A	S	E	T	H	T	A	I	K	G	L	R	Y	Q	K										
	JX898874/ArD149194/SEN/2000	N	A	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	Q	K										
	JX898875/ArD149815/SEN/2000	N	A	R	A	S	K	M	H	T	A	I	R	D	L	R	Y	Q	K										
	AY572535/Gambia2001/GMB/2001	N	A	R	A	S	E	M	H	T	A	I	K	D	-	R	Y	H	K										
	JX898876/ArD156468/SEN/2001	N	A	R	V	F	E	M	H	T	A	I	K	D	L	S	Y	Q	R										
	JX898878/ArD181250/SEN/2005	N	V	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	Q	K										
	JX898879/ArD181676/SEN/2005	N	V	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	Q	K										
	JX898881/ArD181439/SEN/2005	N	V	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	Q	K										
	JX898877/ArD181464/SEN/2005	N	V	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	Q	K										
	JX898880/ArD181564/SEN/2005	N	V	R	A	S	E	M	H	T	A	I	K	D	L	R	Y	O	K										

South America genotype	Genbank accession number	Country	Isolation date
I	MF423373	Brazil	2017
	MF423378		
	MF538785		
	MF423377		
	MF538784		
	MF423374		
	KY885000		
	KY885001		
	MF423376		
	MF538783		
	MF434851		
	MF423375		
	MF538786		
	MF538782		
	KM388815	Venezuela	2007
	KM388818		2006
	KM388814		2005
	KM388816		2010
	KM388817		2004
	KY861728	Brazil	2008
	JF912190		2002
	JF912187		2000
	JF912188		2000
	JF912189		2001
	HM582851	Trinidad-Tobago	2009
	JF912185	Brazil	1992
	JF912182		1984
	JF912180		1981
	JF912184		1987
	JF912179		1980
	JF912186		1994
	JF912183		1984
II	KF907504	Bolivia	1999
	JF912181	Brazil	1983