Yellow fever virus re-emergence and spread in Southeast Brazil, 2016-2019

Running title: Spread of YFV in southeast Brazil

Marta Giovanetti[1,2]^*, Marcos Cesar Lima de Mendonça[1]^, Vagner Fonseca[2,3,5], Maria Angélica Mares-Guia[1], Allison Fabri[1], Joilson Xavier[2], Jaqueline Goes de Jesus[4], Tiago Gräf[4], Cintia Damasceno dos Santos Rodrigues[1], Carolina Cardoso dos Santos[1], Simone Alves Sampaio[1], Flavia Lowen Levy Chalhoub[1], Fernanda de Bruycker Nogueira[1], Julien Theze[5], Alessandro Pecego Martins Romano[6], Daniel Garkauskas Ramos[6], Andre Luiz de Abreu[7], Wanderson Kleber Oliveira[8], Rodrigo Fabiano do Carmo Said[6], Carlos F. Campelo de Alburque[9], Tulio de Oliveira[3], Carlos Augusto Fernandes[10], Shirlei Ferreira Aguiar[10], Alexandre Chieppe[11], Patrícia Carvalho Sequeira[1], Nuno Rodrigues Faria[5], Rivaldo Venâncio Cunha[12,13], Luiz Carlos Junior Alcantara[1,2], Ana Maria Bispo de Filippis[1].

[1]Laboratório de Flavivírus, Instituto Oswaldo Cruz Fiocruz, Rio de Janeiro, Brazil; [2]Laboratório de Genética Celular e Molecular, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [3]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), College of Health Sciences, University of KwaZuluNatal, Durban 4001, South Africa; [4]Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Ministério da Saúde, Salvador-BA; [5]Department of Zoology, University of Oxford, OX1 3PS, UK; [6]Coordenação Geral de Vigilância de Arboviroses (CGARB); [7]Coordenação Geral dos Laboratórios de Saúde Pública/Secretaria de Vigilância em Saúde, Ministério da Saúde, (CGLAB/SVS-MS) Brasília, Distrito Federal, Brazil; [8]Secretaria de Vigilância em Saúde, Ministério da Saúde (SVS-MS), Brasília, Distrito Federal, Brazil; [9]Organização Pan-Americana da Saúde/Organização Mundial da Saúde, Brasília, Distrito Federal, Brazil; [10]Laboratório Central de Saúde Pública Noel Nutels (LACEN-RJ), Rio de Janeiro, Brazil; [11]Superintendência de Vigilância do Estado, Rio de Janeiro, Brazil. [12]Fundação Oswaldo Cruz, Bio-

Manguinhos, Rio de Janeiro, RJ, Brasil; [13]Universidade Federal do Mato Grosso do Sul, Faculdade de Medicina, Campo Grande, MS, Brazil

^Marta Giovanetti and Marcos Cesar Lima de Mendonça contributed equally.

***Correspondence should be addressed to**: giovanetti.marta@gmail.com

Keywords: Yellow fever, outbreak, southeast Brazil, genomic surveillance, outbreak response

ABSTRACT

The recent re-emergence of yellow fever virus (YFV) in Brazil has raised serious concerns due to the virus' rapid dissemination in the southeastern region. To better understand YFV genetic diversity and dynamics during the recent outbreak in southeastern Brazil we generated 18 complete and near-complete genomes from the peak of the epidemic curve from non-human primates (NHPs) and human infected cases across Espírito Santo and Rio de Janeiro states. Genomic sequencing of 18 YFV genomes revealed the estimated timing, source and likely routes of yellow fever virus transmission and dispersion during one of the largest outbreaks ever registered in Brazil. We showed that during the recent epidemic YFV was re-introduced from Minas Gerais to Espírito Santo and Rio de Janeiro states multiple times between 2016 to 2019. The analysis of data from portable sequencing could identify the corridor of spread of YFV. These findings reinforce that continued genomic surveillance strategies can provide information on virus genetic diversity and transmission dynamics that might assist in the understanding arbovirus epidemics.

IMPORTANCE

Arbovirus infections in Brazil including yellow fever, dengue, zika and chikungunya result in considerable morbidity and mortality and are pressing public health concerns. However, our understanding of these outbreaks is hampered by limited availability genomic data. In this study, we

investigated the genetic diversity and spatial distribution of YFV during the current outbreak by analyzing genomic data from areas in southeastern Brazil not covered by other previous studies. To gain insights into the routes of YFV introduction and dispersion, we tracked the virus by sequencing YFV genomes sampled from non-human primates and infected patients from the southeastern region. Our study provides an understanding of how YFV initiates transmission in new Brazilian regions and illustrates that genomics in field can augment traditional approaches to infectious disease surveillance and control.

INTRODUCTION

Yellow fever (YF) is a vector-borne disease that is endemic in tropical areas of Africa and South America (1). The aetiologic agent is the yellow fever virus (YFV), a single-stranded positive sense, RNA virus belonging to the *Flaviviridae* family (2). YFV diversity can be classified into four distinct genotypes, which have been named based on their geographical distribution: East African, West African, South American I, and South American II genotypes (3-6).

In the Americas, YFV transmission can occur via two main epidemiological transmission cycles: the sylvatic (or jungle) and the urban (domestic) cycles. In the sylvatic cycle non-human primates (NHPs) are infected through the bite of mosquito vectors such as *Haemagogus spp.* and *Sabethes spp.* (7, 8). However, in the urban cycle, humans can be infected by *Aedes spp.* mosquitoes biting (9). YFV infection in humans shows a wide spectrum of disease severity including asymptomatic infection, mild illness with dengue-like symptoms, including fever, nausea, vomiting and fatigue, and disease, including fever with jaundice or hemorrhage and death (10).

While eradication is not feasible due to the wildlife reservoir system, large-scale vaccination coverage provides considerable protection against the re-urbanization of YFV transmission (11). However, despite the availability of effective vaccines, YF remains an important public health issue in Africa and South America. In late 2016, a severe re-emergence of YFV epidemic occurred in southeastern Brazil. The epidemic has evolved to become the largest observed in the country in decades, reaching areas close to the Atlantic rainforest (11, 12). YFV 2016-2017 epidemic in Brazil

accounted for 1,412 epizootics, 777 YF human confirmed cases, most of which in southeast Brazil (Minas Gerais n=465; Sao Paulo n=22, Rio de Janeiro n=25; Espírito Santo n=252 confirmed cases), and 261 human deaths (13). Following this epidemic new cases were reported between 2017-2018 and in that period 864 epizootics, 1,376 YF human confirmed cases and 483 human deaths were registered, with the southern states among the most affected by the YFV epidemic (Minas Gerais n=532; Sao Paulo n=377, Rio de Janeiro n=186; Espírito Santo n=6 confirmed cases) (14). The epidemic persisted in 2018-2019 and accounted for 1,883 NHP notified cases (n=20 confirmed NHP cases) and 12 human confirmed cases, including 5 human deaths from the state of São Paulo. Most of the confirmed epizootic cases was registered in the southeastern states (95%) (São Paulo (n=10); Rio de Janeiro (n=8) and Minas Gerais (n=1) (13-15).

Although there is currently no evidence that urban transmission has occurred, the outbreak affected areas highly infested by *Ae. aegypti* and *Ae. Albopictus* where yellow fever vaccination was recently introduced in routinely of immunization program. This raises concern that, for the first time in decades, there might be high risk of YFV urban transmission in Brazil (16). New surveillance and analytical approaches are therefore needed to monitor this threat.

Even so, there is limited information from genomic surveillance studies about the genomic epidemiology and the dissemination dynamics of 2016-2019 YFV circulating in Southeast Brazil. Previous studies have shown the spatial and evolutionary dynamics of the current YFV outbreak in different southeastern states, (11) and shed light regarding the possible co-circulation of distinct YFV lineages (17). Nevertheless, there is still limited information about the genomic epidemiology of YFV circulating in Espírito Santo and Rio de Janeiro states from genomic surveillance studies, and this impairs our understanding of the virus re-introduction, establishment and dissemination in those regions. Thus, to better understand the re-emergence of the recent YFV epidemic in those regions, we analyzed a larger and updated dataset of recently released data of the YFV 2016-2019 epidemic in Brazil, including 18 newly generated complete genomes from areas not covered by other previous studies from human and NHPs from the Southeast states of Espírito Santo and Rio de Janeiro.

RESULTS

Molecular diagnostics and genome sequencing from clinical samples

Liver, spleen, kidney and blood samples from 14 NHPs and liver and serum samples from 4 human infected cases collected from areas not covered by other previous studies in Rio de Janeiro and Espírito Santo states, Southeast Brazil, between January 2017 and April 2018, were tested for YFV RNA using the RT-qPCR assay (18, 19) at the Flavivirus Laboratory at FIOCRUZ Rio de Janeiro (LABFLA/FIOCRUZ).

Most confirmed cases in NHPs were from animals of the *Alouatta* genus (42.9%; 6 of 14), followed by *Callithrix* (35.7%; 5 of 14), *Sapajus* (7.1%; 1 of 14) and *Leontopithecus rosalia* (14.3%; 2 of 14). PCR cycle threshold (Ct) values were on average 12.23 (range: 7.2 to 22.4) (**Table 1**).

The median age of human patients was 38 years (range: 16 to 65 years). A total of 75% of the affected subjects lived in rural areas (**Table 1**). Only one subject lived in urban areas, with a history of travel to rural areas. To investigate the source and transmission of YFV and the genetic diversity of the virus circulating in human and NHPs across Rio de Janeiro and Espírito Santo states, we used the MinION handheld nanopore sequencer to generate 18 complete and near complete genomic sequences (average coverage = 89.9%; **Table 2**) using a previously described MinION sequencing protocol (11, 20) that allowed the rapid data generation through a fast sample preparation and library construction (1 day) as an interesting approach to get rapid critical information (such as lineage identification and pathogen transmission dynamics) useful for the surveillance services and decision-makers.

YF samples sequenced in this study were geographically widespread across 6 municipalities of Rio de Janeiro and 7 municipalities of Espírito Santo (**Figure 1A**).

Figure 1 panel B shows the number of YFV confirmed cases in the Espirito Santo and Rio de Janeiro states respectively. Epidemiological data revealed two distinct YFV epidemic waves. The first epidemic wave (*wave 1*) is represented by the YFV cases mainly registered in the Espírito

Santo state, during the first semester of 2017 (January to April, *n*= 252 cases), although some sporadic cases were reported in the following year (**Figure 1B**). The second wave (*wave 2*), is represented by YFV cases registered in Rio de Janeiro state during first semester of 2018 (February to May; *n*=220 cases) (**Figure 1B**). Although majority of cases in Rio de Janeiro occurred between February and March 2018, there was also a re-emergence of YFV in that state detected around March 2017, during epidemic wave 1 that mainly affected Espírito Santo state.

Genetic history of YFV in Southeastern Brazil

To investigate the phylogenetic relationship of YFV strains circulating in the southeastern states of Espírito Santo and Rio de Janeiro we estimated a maximum likelihood (ML) phylogenetic tree for a dataset of 181 reference sequences comprising the four YFV lineages. Our ML phylogeny revealed that, as suspected, the newly generated YFV sequences belong to the South American I (SAI) lineage with high statistical support (bootstrap = 100%), clustering with other Brazilian isolates from the 2016-2019 epidemic (**Figure 2**).

Subsequently, to investigate the dynamics of the YFV epidemic within the Southeast region, genetic analyses were conducted on a second dataset (dataset 2, *n* = 137), including recently published sequences from the YFV 2016-2019 epidemic in Brazil, belonging to the SA1 lineage. The timescale of our phylogenetic estimates was consistent with recently studies (17, 21, 22) and confirmed the presence of two distinct lineages circulating in the current YFV epidemic, named hereafter as SA1 lineage 1 and SA1 lineage 2 (**Figure 3**). The SA1 lineage 1 comprises sequences from the northern and eastern regions of Minas Gerais, Bahia, Espírito Santo and Rio de Janeiro states and the time of the most recent common ancestor (TMRCA) of this lineage was dated back to September 2016 (95% BCI: July to November 2016). The SA1 lineage 2 comprises sequences from the southern municipalities of Minas Gerais state and sequences from the southeastern state of Sao Paulo and the TMRCA of this lineage was dated back around July 2016 (95% BCI: June to December 2016) (**Figure 3**). Our time-scaled phylogeny showed that the sequences generated in this study clustered together with high support (pp=90 %) within SA1 lineage 1 (**Figure 3**).

In order to understand the transmission and the spatio-temporal evolution of the SA1 lineage 1, we analysed a subset of 80 (Dataset 3) sequences representing all the available sequences from this lineage (**Figure 4**). We performed a regression of genetic divergence from root to tip against sampling dates that confirmed sufficient temporal signal (r^2=0.70) in this dataset. A time-scaled phylogenetic analysis using a Bayesian Markov Chain Monte Carlo (MCMC) framework (23) was then performed to investigate the time of introduction of the YFV into the Espírito Santo and Rio de Janeiro states (**Figure 5 A**). **Figure 5A** shows a zoom of our Bayesian time-scaled phylogeny highlighting the SA1 lineage 1 comprising the 2017-2019 YFV strains from Minas Gerais, Bahia, Espírito Santo and Rio de Janeiro states. Our analysis showed that samples from Espírito Santo were intermixed with sequences from Rio de Janeiro. This suggests that the YFV epidemic in Espírito Santo and Rio de Janeiro was not caused by a single introduction event, as observed in Sao Paulo (17, 21), but resulted from multiples introductions over time.

We next used a continuous diffusion model to investigate how the SA1 lineage 1 has been spreading over space and time. We found evidence that YFV disseminated through southeastern Brazilian states using two distinct paths with an average dispersal rate of 0.12 km/day (95% HPD: 0.09 – 0.14 km/day). From the northern region of Minas Gerais state, YFV spread to the south region of Bahia state around January 2017 (95% BCI: December 2016 to February 2017) (**Figure 5**), and from the eastern region of Minas Gerais state YFV moved towards Espírito Santo state, (pp=0.99) with introductions estimated around January 2017 (95% BCI: November 2016 to January 2017) (**Figure 5 Panels A, B**). Since its introduction in the Espírito Santo state, the virus spread through the neighboring state (**Figure 5**). Our analyses revealed that YFV was likely introduced in Rio de Janeiro state several times between January (95% BCI: December 2016 to February 2017) and March 2017 (February 2017 to May 2017), spreading southward from the border with Espírito Santo state and reaching Angra dos Reis municipality, which is located in the southern region of Rio de Janeiro. Our data further suggests that after its first introduction in Rio de Janeiro the virus persisted until 2019, as indicated by the isolate MK533792 sampled, in the

188 municipality of Casimiro de Abreu, in January 2019 (12) (**Figure 5 Panel A**), reinforcing the need
189 for maintaining continuous surveillance and high vaccination coverage in the Southeastern region.
190
191 **DISCUSSION**
192 In this study, we generated and analysed 18 new YFV complete and near-complete genomic
193 sequences from samples from humans and non-human primates collected in several municipalities
194 not covered by other previous studies in Espírito Santo and Rio de Janeiro states, in 2017-2018.
195 Although previous studies have already shown the spatial and evolutionary dynamics of the current
196 YFV outbreak in Brazil (11, 12, 17, 21), the shortage of genomic data from Espirito Santo and Rio
197 de Janeiro states hampered the ability to shed light on the re-emergence and establishment of the
198 YFV transmission in those regions. Try to determine in a large scale the corridor of spread of YFV
199 and the geographic hot spots is key to predicting and preventing other possible spillover events. The
200 generated genomic data provides a more detailed understanding of the introduction and progression
201 of YFV SA1 lineage 1 and reveals the timing, source and likely routes of yellow fever virus
202 transmission and dispersion during the largest outbreak in Brazil in decades.
203 According to the Ministry of Health epidemiological bulletin, the YFV re-emergence in the
204 states of Espírito Santo and Rio de Janeiro was confirmed in these states in January and February
205 2017 respectively (13-15).
206 Our estimates indicated that YFV strains from the epidemic first emerged in Espírito Santo
207 state from Minas Gerais around January 2017 (95% BCI: November 2016 to January 2017), which
208 is consistent with epidemiological data (13-15). From the state of Espírito Santo, YFV spread
209 southwards to the great metropolitan area of Rio de Janeiro state. Moreover, our data indicated that
210 the circulation of YFV in Rio de Janeiro may have resulted from multiple and independent
211 introductions events from Espírito Santo state, highlighting a complex dispersion dynamic of the
212 current YFV outbreak in Brazil which occurred between January (95% BCI: December 2016 to
213 February 2017) and March 2017 (February 2017 to May 2017). Our data further suggest that after
214 its first introduction in Rio de Janeiro the virus persisted until 2019, as indicated by the isolate

MK533792 sampled in the municipality of Casimiro de Abreu in January 2019 (12). This estimation suggests the YFV might have been persisted in Rio de Janeiro state for approximately 24 months. This suggests that Rio de Janeiro state possibly possesses the ecological conditions to maintain YFV outside the period of transmission (Dec to May) (12). Ultimately, given the abundance of sylvatic competent vectors (12) and non-human primates (21, 22), this data could indicate that there is some potential for the establishment of an enzootic transmission cycle of yellow fever in Mata Atlantica.

Epidemiological data also indicated two distinct YFV epidemic waves (13, 14). The first epidemic wave is represented by the YFV cases mainly registered in Minas Gerais and Espírito Santo state during the first semester of 2017, while the second wave is represented by the YFV cases registered in Rio de Janeiro state during first semester of 2018. Transmission of YFV in areas with susceptible NHPs species typically occurs in time periods characterized by environmental conditions suitable to support higher mosquito abundance (12, 24).

As previously suggested (17), we found evidence regarding the circulation of two distinct YFV lineage, that might have been spread to distinct evolutionary and diffusion rates. Using YFV genetic data, we estimate that the YFV SA1 lineage 1 spread at rate of 0.12 km/day (95% HPD: 0.09 – 0.14 km/day), that is slightly lower than previously estimates (11, 17), this might be due to the larger dataset analyzed in this study, that might explain differences in the rate of YFV spread among different areas as well as different lineage.

These findings reinforce that continued genomic surveillance strategies are needed to assist in the monitoring and understanding of arbovirus epidemics, which might help to attenuate the public health impacts of infectious diseases. In the present research article we aimed to provided genomic information and reinforce the idea of the use of epidemiological and genomic data, generated by a portable, ease of setup sequencing system, as an approach to get rapid critical information (such as lineage identification and pathogen transmission dynamics) that could be used by the surveillance services and decision-makers.

In this study we also demonstrate that by analyzing heterochronous datasets with samples collected in different time points and/or locations, phylodynamics becomes a powerful tool to prevent and identify the viral lineage movement and to describe trends in epidemic spread (11, 25, 26).

Continued surveillance in human and non-human primates (NHP) in non-epidemic periods in the southeast region will be important in order to quantify the risk of new outbreaks and the establishment of new YFV transmission cycles in the region. In conclusion, our study shows that genomic data generated by portable sequencing technology can be employed to assist public health services in monitoring and understanding the diversity of circulating mosquito-borne viruses.

MATERIALS AND METHODS

Sample collection

Human and non-human primate samples were collected, under the guidelines of a national strategy of YF surveillance, for molecular diagnostics by the Flavivirus Laboratory (LABFLA) at Oswaldo Cruz Foundation (Fiocruz) in Rio de Janeiro, Brazil, which is a Brazilian Ministry of Health Regional Reference Laboratory for arboviruses. The majority of samples were linked to a digital record that collated epidemiological and clinical data such as date of sample collection, municipality of residence, neighborhood of residence, demographic characteristics (age and sex) and date of onset of clinical symptoms.

Ethical statement

The project was supported by the Pan American World Health Organization (PAHO) and the Brazilian Ministry of Health (MoH) as part of the arboviral genomic surveillance efforts within the terms of Resolution 510/2016 of CONEP (Comissão Nacional de Ética em Pesquisa, Ministério da Saúde; National Ethical Committee for Research, Ministry of Health). The

268 diagnostic of YFV infection at LABFLA was approved by the Ethics Committee of the Oswaldo
269 Cruz Institute CAAE90249218.6.1001.54248.
270
271 ***RT-qPCR***
272 Total RNA was extracted from tissue and serum samples using MagMAX™ Pathogen RNA/DNA
273 kit (Life Technologies™, Carlsbad CA, USA) in accordance with the manufacturer's instructions.
274 Viral RNA was detected using two previously published RT-qPCR techniques (18, 19).
275
276 ***cDNA synthesis and whole genome nanopore sequencing***
277 Sequencing was attempted on the 18 selected RT-PCR positive samples regardless of Ct value as
278 previously described (11, 20, 26). All positive samples were submitted to a cDNA synthesis
279 protocol (11, 20) using ProtoScript II First Strand cDNA Synthesis Kit. Then, a multiplex tiling
280 PCR was attempted using the previously published YFV primer scheme and 30 cycles of PCR using
281 Q5 High-Fidelity DNA polymerase (NEB) as previously described (20). Amplicons were purified
282 using 1x AMPure XP Beads (Beckman Coulter) and cleaned-up PCR products concentrations were
283 measured using Qubit™ dsDNA HS Assay Kit on a Qubit 3.0 fluorimeter (ThermoFisher). DNA
284 library preparation was performed using the Ligation Sequencing Kit (Oxford Nanopore
285 Technologies) and the Native Barcoding Kit (NBD103, Oxford Nanopore Technologies, Oxford,
286 UK). Sequencing library was generated from the barcoded products using the Genomic DNA
287 Sequencing Kit SQK-MAP007/SQK-LSK208 (Oxford Nanopore Technologies). Sequencing
288 library was loaded onto a R9.4 flow cell (Oxford Nanopore Technologies).
289
290 ***Generation of consensus sequences***
291 Consensus sequences for each barcoded sample were generated following a previously published
292 approach (20). Briefly, raw files were basecalled using Albacore, demultiplexed and trimmed using
293 Porechop, and then mapped with *bwa* to a reference genome (GenBank accession number
294 JF912190). Nanopolish variant calling was applied to the assembly to detect single nucleotide

295 variants to the reference genome. Consensus sequences were generated; non-overlapped primer
296 binding sites, and sites for which coverage was <20X were replaced with ambiguity code N.
297 Sequencing statistics can be found in **Table 1**.
298
299 ***Collation of YFV complete genome datasets***
300 Genotyping was first conducted using the phylogenetic Yellow fever typing tool available at
301 http://www.krisp.org.za/tools.php. The genome sequences generated here were combined with a
302 dataset comprising previously published genomes from the 2016-2019 YFV epidemic in Brazil
303 (11,12, 17, 21, 22). Two complete or near-complete YFV genome datasets were generated. Dataset
304 1 (*n* = 199) comprised the data reported in this study (*n* = 18) plus (*n* = 181) complete or almost
305 complete YFV genomic sequences (>10,000 bp), retrieved from NCBI in June 2019 and covering
306 all four existing genotypes. Subsequently, to investigate the dynamic of the YFV infection within
307 the Southeast region, genetic analyses were conducted on a smaller dataset (dataset 2) including a
308 larger and updated dataset of recently released data of the YFV 2016-2019 epidemic in Brazil
309 belonging to the SA1 lineage (*n* = 137). Thus, to understand the transmission and the spatio-
310 temporal evolution of the YFV SA1 lineage 1, from this dataset, we generated a subset (dataset 3)
311 that included all identified sequences from that lineage (*n*= 80). Maximum likelihood (ML)
312 phylogenetic trees were estimated using RAxML (27) under a GTR + Γ_4 nucleotide substitution
313 model. Statistical support for phylogenetic nodes was estimated using a ML bootstrap approach
314 with 1000 replicates.
315 In order to investigate the temporal signal in our YFV datasets 2 and 3 we regressed root-to-tip
316 genetic distances from this ML tree against sample collection dates using TempEst v 1.5.1
317 (http://tree.bio.ed.ac.uk) (28).
318
319 ***Dated phylogenetics***
320 To estimate time-calibrated phylogenies dated from time-stamped genome data, we conducted
321 phylogenetic analysis using a Bayesian software package (23). Here we used the GTR + Γ_4

nucleotide substitution model and Bayesian Skygrid tree prior (29) with an uncorrelated relaxed clock with a lognormal distribution (30). We employed a stringent model selection analysis using both path-sampling (PS) and stepping-stone (SS) procedures to estimate the most appropriate molecular clock model for the Bayesian phylogenetic analysis [38]. We tested: a) the strict molecular clock model, which assumes a single rate across all phylogeny branches, and b) the more flexible uncorrelated relaxed molecular clock model with a lognormal rate distribution (UCLN) [39]. Both SS and PS estimators indicated the strict molecular clock (Bayes Factor = 4.3) as the best fitted model to the dataset under analysis. Analyses were run in duplicate in BEASTv.1.10.4 (23) for 50 million MCMC steps, sampling parameters and trees every 5000th step. A non-informative continuous time Markov chain reference prior on the molecular clock rate was used (31). Convergence of MCMC chains was checked using Tracer v.1.7.1 (32). Maximum clade trees were summarized using TreeAnnotator after discarding 10% as burn-in.

Phylogeographic analyses

To investigate the spread of YFV in Southeast Brazil, we analysed in more detail the SA1 Lineage 1 that includes the n=80 sequences (**Figure 4**). We used a skygrid coalescent tree prior (33) and a continuous phylogeographic model that uses a relaxed random walk to model the spatial diffusion of lineages. Dispersal velocity variation among lineages was modelled using a Cauchy distribution (34, 35). Virus diffusion through time and space was summarized using 1000 phylogenies sampled at regular intervals from the posterior distribution (after exclusion of burn-in). Sampling location of each geo-referenced YFV sequences from Espírito Santo and Rio de Janeiro state are listed in **Supplementary table 1**. Georeferenced and time-stamped sequences were analysed in BEAST v.1.10.4 (23) using the BEAGLE library (36) to enhance computational speed.

Data availability

New sequences were deposited in GenBank under accession numbers MK882599 to MK882604, MK882607 to MK882613, MK882615, MK882617 to MK882619, and MK882621 (see Table 2).

ACKNOWLEDGMENTS

The authors thank all personnel from Health Surveillance System from Rio de Janeiro that coordinated surveillance and helped with data collection and assembly. The authors also thank Ronaldo Lapa e Solange Regina Conceição from the Instituto Oswaldo Cruz (Fiocruz – RJ) for their support. This work was supported by the ZiBRA2 project supported by the Brazilian Ministry of Health (SVS-MS) and the Pan American Organization (OPAS) and founded by Decit/SCTIE/MoH and CNPq (440685/2016-8 and 440856/2016-7); by CAPES (88887.130716/2016-00, 88881.130825/2016-00 and 88887.130823/2016-00); by EU's Horizon 2020 Programme through ZIKAlliance (PRES-005-FEX-17-4-2-33). The Flavivirus Laboratory activities was also supported by Faperj (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro) under the grant no. E-26/2002.930/2016, by the International Development Research Centre (IDRC) Canada over the grant 108411-001 and Horizon 2020 through ZikaPlan and ZikAction under the grant's agreement No. 734584 and 734857. MG is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ.

M.G., M.C.L.M., V.F., L.C.J.A., and A.M.B.F. conceived and designed the study. M.G., M.C.L.M., V.F., M.A.M.G., A.F., J.X., J.G.J., C.D.S.R., C.C.S., P.C.S., S.A.S., F.L.LC., F.B.N., and J.T. performed investigations. M.G., M.C.L.M., V.F., T.G., J.T., N.R.F., A.P.M.R., D.G.R., A.L.A., W.K.O., R.F.C.S., C.F.C.A., T.O. C.A.F., S.F.A., R.V.C., A.C., L.C.J.A., and A.M.B.F. curated the data. M.G., V.F., T.G., J.T., N.R.F., L.C.J.A., and A.M.B.F. performed formal analysis. M.G., M.C.L.M., V.F., L.C.J.A., and A.M.B.F. wrote the original draft of the paper, and M.G., M.C.L.M., T.O., P.C.S., A.P.M.R., D.G.R., R.V.C., N.R.F., R.V.C., and A.M.B.F. revised the paper. A.P.M.R., D.G.R., A.L.A., W.K.O., R.F.C.S., C.F.C.A., C.A.F., S.F.A., R.V.C., A.C., and A.M.B.F. provided resources.

376 The authors declare no competing interests.

377

378 **FIGURE LEGENDS**

379

380 **Figure 1. Spatial and temporal distribution of YF cases from Espírito Santo and Rio de Janeiro states**
381 **during 2017 and 2019.**

382 **A**. Map of the states of Espírito Santo (ES) and Rio de Janeiro (RJ), located in south-eastern region of Brazil,
383 and its municipalities. Circles indicate where samples from this study were collected. **B.** Time series of
384 human (H) and Non-human primate YFV cases in ES and RJ states confirmed by serology, reverse
385 transcription quantitative PCR (RT-qPCR), or virus isolation. Below, the dates of sample collection of the
386 virus genomes generated in this study are shown in grey bars.

387

388 **Figure 2**. Molecular phylogenetics of the Brazilian YFV epidemic. Maximum likelihood phylogeny of
389 complete YFV genomes showing the outbreak clade (gray triangle) within the South American I (SA1)
390 genotype. The scale bar is in units of substitutions per site (s/s).

391

392 **Figure 3. Time scaled phylogenetic tree of the current YF epidemic in Brazil.**

393 Molecular clock phylogeny obtained by combining the 18 new YFV complete genomic generated here
394 (starred tips), plus public available data (n=137) of the YFV 2016-2019 epidemic in Brazil (11; 12; 17; 21-
395 22). Numbers in nodes represent clade posterior probability >0.90. Branch colours represent different
396 sampling locations.

397

398 **Figure 4**. Molecular clock phylogeny including the clade comprising the new isolates plus all the YFV
399 strains from the 2017-2019 outbreak belonging to the SA1 lineage 1 clade. Numbers along branches
400 represent clade posterior probability >0.90. Colours represent different locations.

401

402 **Figure 5. Spatio-temporal dynamics of the YFV SA1 lineage 1.**

403 **A**. Molecular clock phylogeny including the clade comprising the 2017-2019 YFV strains from Minas
404 Gerais, Bahia, Espírito Santo and Rio de Janeiro states belonging to the SA1 lineage 1. Numbers along

branches represent clade posterior probability >0.90. YFV isolates from Casimiro de Abreu, sampled in January 2019 is highlighted in red. Colours represent different locations. **B**. Reconstructed spatiotemporal continuous diffusion of the YFV SA1 lineage 1 outbreak clade. Phylogenetic branches are mapped in space according to the location of phylogenetic nodes (circles). Lines show the cross-state movement of the virus from Minas Gerais followed by movement to the states of Espírito Santo and Rio de Janeiro. Shaded regions show 95% credible regions of internal nodes.

REFERENCES

1. Barrett AD, Monath TP. 2003. Epidemiology and ecology of yellow fever virus. Adv Virus Res 61:291–315.
2. Monath TP, Vasconcelos PF. 2015. Yellow fever. J Clin Virol 64:160–173.
3. Bryant JE, Holmes EC, Barrett AD. 2007. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathog 3:75-77.
4. Mutebi JP, Wang H, Li L, Bryant JE, Barrett AD. 2001. Phylogenetic and evolutionary relationships among yellow fever virus isolates in Africa. J Virol 75:6999–7008.
5. Nunes MR, Palacios G, Cardoso JF, Martins LC, Sousa EC Jr, de Lima CP, Medeiros DB, Savji N, Desai A, Rodrigues SG, Carvalho VL, Lipkin WI, Vasconcelos PF. 2012. Genomic and phylogenetic characterization of Brazilian yellow fever virus strains. J Virol 86:13263-71.
6. Chang GJ, Cropp BC, Kinney RM, Trent DW, Gubler DJ. 1995. Nucleotide sequence variation of the envelope protein gene identifies two distinct genotypes of yellow fever virus. J Virol 69:5773-80.
7. Dégallier N, Travassos da Rosa AP, Hervé JP, Travassos da Rosa JFS, Vasconcelos PFC, Silva CJM. 1992. A comparative study of yellow fever in Africa and South America. J Braz Assoc Advanc Sci 44:143–51.

8. Mondet B, Travassos da Rosa APA, Vasconcelos PFC. 1996. The risk of urban yellow fever outbreaks in Brazil by dengue vectors Aedes aegypti and Aedes albopictus. Bull Soc Pathol Exot 89:107–13.
9. Souza RP de, Petrella S, Coimbra TLM, Maeda AY, Rocco IM, Bisordi I, Silveira VR, Pereira LE, Suzuki A, Silva SJDS. 2011. Isolation of yellow fever virus (YFV) from naturally infected Haemagogus (Conopostegus) leucocelaenus (diptera, cukicudae) in São Paulo State, Brazil, 2009. Rev Inst Med Trop São Paulo 53:133–139.
10. Ribeiro AF, Cavalin RF, Abdul Hamid Suleiman JM, Alves da Costa J, Januaria de Vasconcelos M, Sant'Ana Málaque CM, Sztajnbok J. Yellow Fever: Factors Associated with Death in a Hospital of Reference in Infectious Diseases, São Paulo, Brazil, 2018. 2019. Am J Trop Med Hyg 20:05- 28.
11. Faria NR, Kraemer MUG, Hill SC, Goes de Jesus J, Aguiar RS, Iani FCM, Xavier J, Quick J, du Plessis L, Dellicour S, Thézé J, Carvalho RDO, Baele G, Wu CH, Silveira PP, Arruda MB, Pereira MA, Pereira GC, Lourenço J, Obolski U, Abade L, Vasylyeva TI, Giovanetti M, Yi D, Weiss DJ, Wint GRW, Shearer FM, Funk S, Nikolay B, Fonseca V, Adelino TER, Oliveira MAA, Silva MVF, Sacchetto L, Figueiredo PO, Rezende IM, Mello EM, Said RFC, Santos DA, Ferraz ML, Brito MG, Santana LF, Menezes MT, Brindeiro RM, Tanuri A, Dos Santos FCP, Cunha MS, Nogueira JS, Rocco IM, da Costa AC, Komninakis SCV, Azevedo V, Chieppe AO, Araujo ESM, Mendonça MCL, Dos Santos CC, Dos Santos CD, Mares-Guia AM, Nogueira RMR, Sequeira PC, Abreu RG, Garcia MHO, Abreu AL, Okumoto O, Kroon EG, de Albuquerque CFC, Lewandowski K, Pullan ST, Carroll M, de Oliveira T, Sabino EC, Souza RP, Suchard MA, Lemey P, Trindade GS, Drumond BP, Filippis AMB, Loman NJ, Cauchemez S, Alcantara LCJ, Pybus OG. 2018. Genomic and epidemiological monitoring of yellow fever virus transmission potential 2018. Science 361:894–899.
12. Abreu FVS, Delatorre E, Araújo Cunha A, Ferreira-de-Brito A, de Castro MG, Ribeiro I, Dias N, Vargas, Paixão W, Ribeiro MS, Meneguete P, Bonaldo MC, Bello G, Lourenço-de-

460 Oliveira R. 2019. Combination of surveillance tools reveals that Yellow Fever virus can
461 remain in the same Atlantic Forest area at least for three transmission seasons. Memórias do
462 Instituto Oswaldo Cruz 114:19-76.
463 13. Brazilian Ministry of Health 2017. Epidemiological Update Yellow fever.
464 (http://portalarquivos2.saude.gov.br/images/pdf/2017/setembro/06/2017_027.pdf).
465 14. Brazilian Ministry of Health 2017 2018. Epidemiological Update Yellow fever.
466 (http://portalarquivos2.saude.gov.br/images/pdf/2018/outubro/08/Informe-FA.pdf).
467 15. Secretaria de Vigilância em Saúde (2019). Monitoramento do Período Sazonal da Febre
468 Amarela Brasil – 2018/2019. Brazil: Ministério da Saúde.
469 16. Shearer FM. 2017. Lancet Infect. Dis 17:1209–1217.

470 17. Delatorre E, de Abreu FVS, Ribeiro IP, Gómez MM, Dos Santos AAC, Ferreira-de-Brito A,
471 Neves MSAS, Bonelly I, de Miranda RM, Furtado ND, Raphael LMS, da Silva LFF, de
472 Castro MG, Ramos DG, Romano APM, Kallás EG, Vicente ACP, Bello G, Lourenço-de-
473 Oliveira R, Bonaldo MC. 2019. Distinct YFV Lineages Co-circulated in the Central-
474 Western and Southeastern Brazilian Regions From 2015 to 2018. Front Microbiol 24;10-
475 1079.
476 18. Drosten C, Göttig S, Schilling S, Asper M, Panning M, Schmitz H, Günther S. 2002. Rapid
477 detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-
478 Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus
479 by real-time reverse transcription-PCR. J Clin Microbiol 40:2323–2330.

480 19. Domingo C, Patel P, Yillah J, Weidmann M, Méndez JA, Nakouné ER, Niedrig M. 2012.
481 Advanced yellow fever virus genome detection in point-of-care facilities and reference
482 laboratories. J Clin Microbiol 50:4054–4060.

483 20. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G,
484 Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG,
485 Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC Jr, Sabino

EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 12:1261–1276.

21. de Souza RP, Hill SC, Thézé J, Claro I, Aguiar RS, Dellicour D, Abade L, Santos FCP, Cunha MS, Nogueira JS, Salles FCS, Rocco IM, Maeda AY, Vasami FGS, du Plessis L, Silveira PP, Giovanetti M, de Goes J, Quick J, Fernandes NCCA, Guerra J, Réssio MRA, Cirqueira CS, Iglezias SD, Delgado JD, Macedo FLL, Timenetsky MCST, de Paula R, Spinola R, Deus JT, Mucci LF, Tubaki RM, Menezes RMT, Ramos PL, Abreu AL, Cruz LN, Loman N, Bispo A, Pybus OG, Alcantara LVJ, Sabino EC, Faria NR. 2019. Genomic Surveillance of Yellow Fever Virus Epidemic Waves in São Paulo, Brazil, 2017 – 2018. bioRxiv 645341; doi: https://doi.org/10.1101/645341.
22. Cunha MS, da Costa AC, de Azevedo Fernandes NCC, Guerra JM, Dos Santos FCP, Nogueira JS, D'Agostino LG, Komninakis SV, Witkin SS, Ressio RA, Maeda AY, Vasami FGS, Kaigawa UMA, de Azevedo LS, de Souza Facioli PA, Macedo FLL, Sabino EC, Leal É, de Souza RP. Epizootics due to Yellow Fever Virus in São Paulo State, Brazil: viral dissemination to new areas (2016-2017). Sci Rep. 2019 Apr 2;9(1):5474.
23. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:1-016.
24. Possas C, Lourenço-de-oliveira R, Tauil PL, Pinheiro FDP, Pissinatti A, da Cunha RV, et al. Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. Mem Inst Oswaldo Cruz. 2018; 113(10): e180278.
25. Gardy JL. Loman NJ. 2017. Towards a genomics-informed, real-time, global pathogen surveillance system. Nature Reviews Genetics. 217; 1:2-256.
26. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, Kraemer MUG, Hill SC, Black A, da Costa AC, Franco LC, Silva SP, Wu CH, Raghwani J, Cauchemez S, du Plessis L, Verotti MP, de Oliveira WK, Carmo EH, Coelho GE, Santelli ACFS, Vinhal LC, Henriques CM, Simpson JT, Loose M, Andersen KG, Grubaugh ND, Somasekar S, Chiu

513 CY, Muñoz-Medina JE, Gonzalez-Bonilla CR, Arias CF, Lewis-Ximenez LL, Baylis SA,
514 Chieppe AO, Aguiar SF, Fernandes CA, Lemos PS, Nascimento BLS, Monteiro HAO,
515 Siqueira IC, de Queiroz MG, de Souza TR, Bezerra JF, Lemos MR, Pereira GF, Loudal D,
516 Moura LC, Dhalia R, França RF, Magalhães T, Marques ET Jr, Jaenisch T, Wallau GL, de
517 Lima MC, Nascimento V, de Cerqueira EM, de Lima MM, Mascarenhas DL, Neto JPM,
518 Levin AS, Tozetto-Mendoza TR, Fonseca SN, Mendes-Correa MC, Milagres FP, Segurado
519 A, Holmes EC, Rambaut A, Bedford T, Nunes MRT, Sabino EC, Alcantara LCJ, Loman NJ,
520 Pybus OG. 2017. Establishment and cryptic transmission of Zika virus in Brazil and the
521 Americas. Nature. 546:406-410.

522 27. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
523 large phylogenies. Bioinformatics 30:1312–1313.

524 28. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure
525 of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2:11-70.

526 29. Shapiro B, Rambaut A, Drummond AJ. 2006. Choosing appropriate substitution models for
527 the phylogenetic analysis of protein-coding sequences. Molecular biology and evolution
528 23:7-9.

529 30. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating
530 with confidence. PLoS biology 54:4-88.

531 31. Ferreira MAR, Suchard MA. 2008. Bayesian analysis of elapsed times in continuous-time
532 Markov chains. Can J Stat 78:355-368.

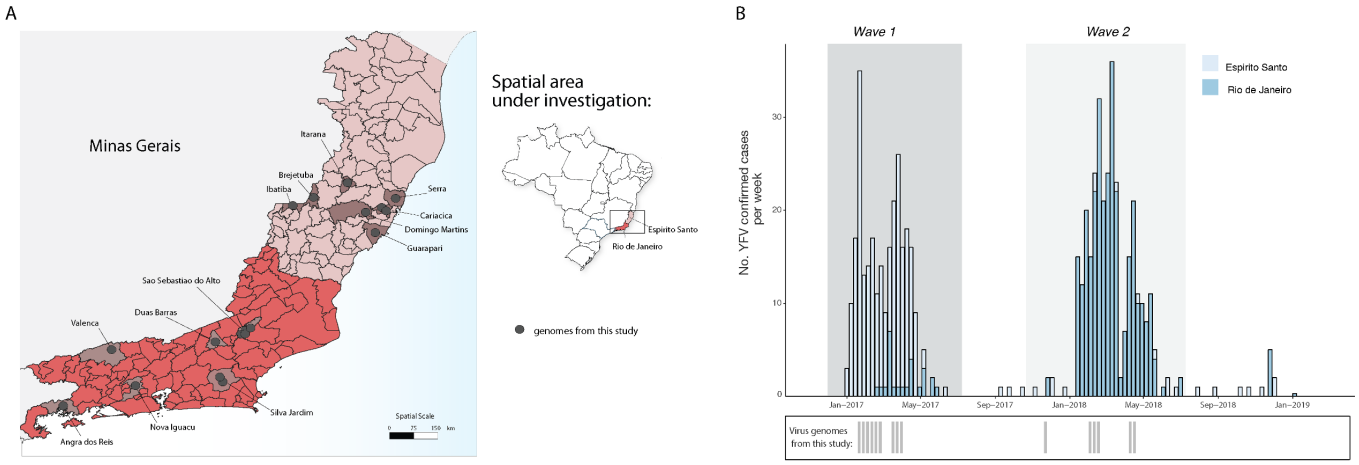
533 32. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in
534 Bayesian phylogenetics using Tracer 1.7. Systematic Biology 2:100-250.

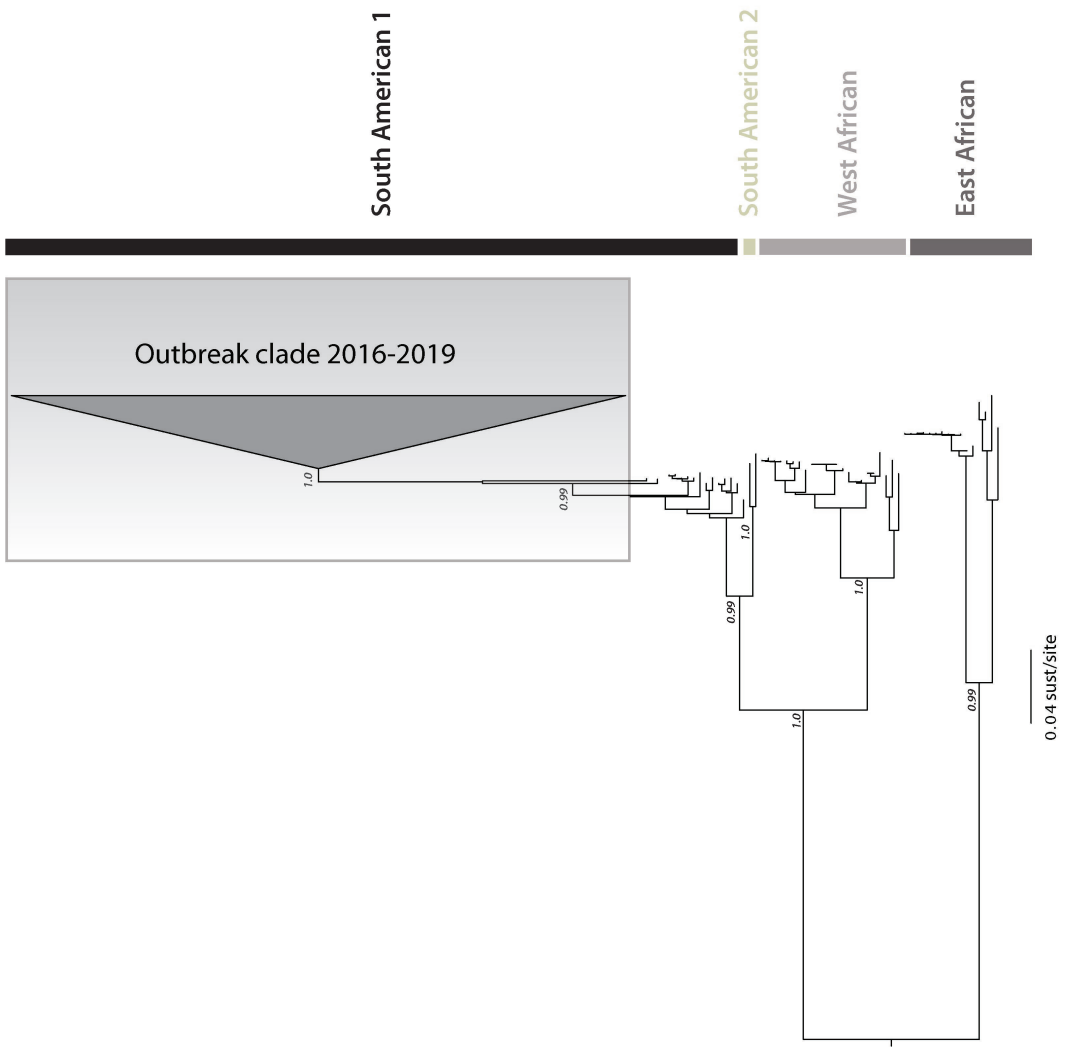
535 33. Shapiro B, Rambaut A, Drummond AJ. 2006. Choosing appropriate substitution models for
536 the phylogenetic analysis of protein-coding sequences. Molecular biology and evolution
537 23:7-9.

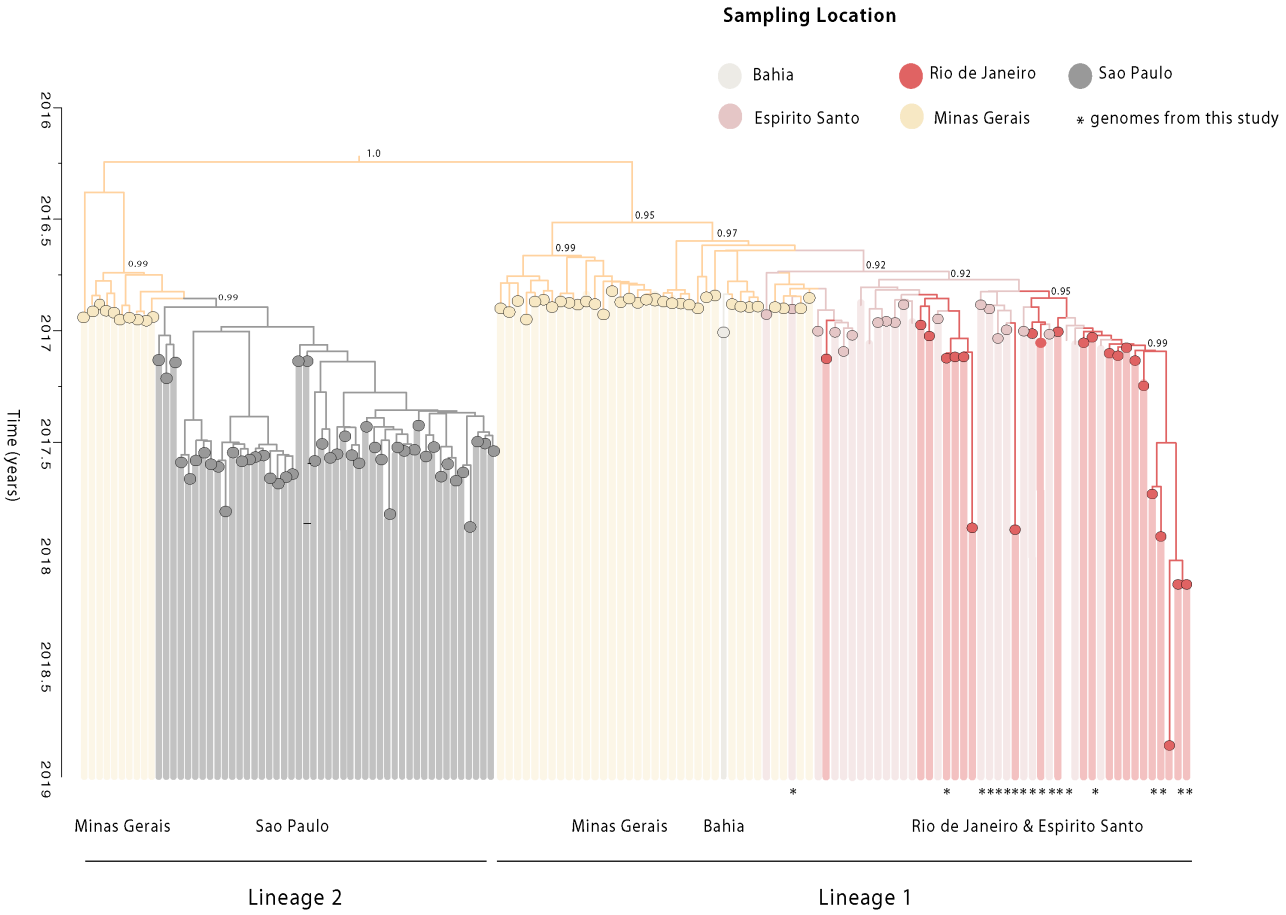
538 34. Lemey P, Rambaut A, Welch JJ, Suchard MA. 2010. Phylogeography Takes a Relaxed
539 Random Walk in Continuous Space and Time. Mol Biol Evol 27:1877–1885.

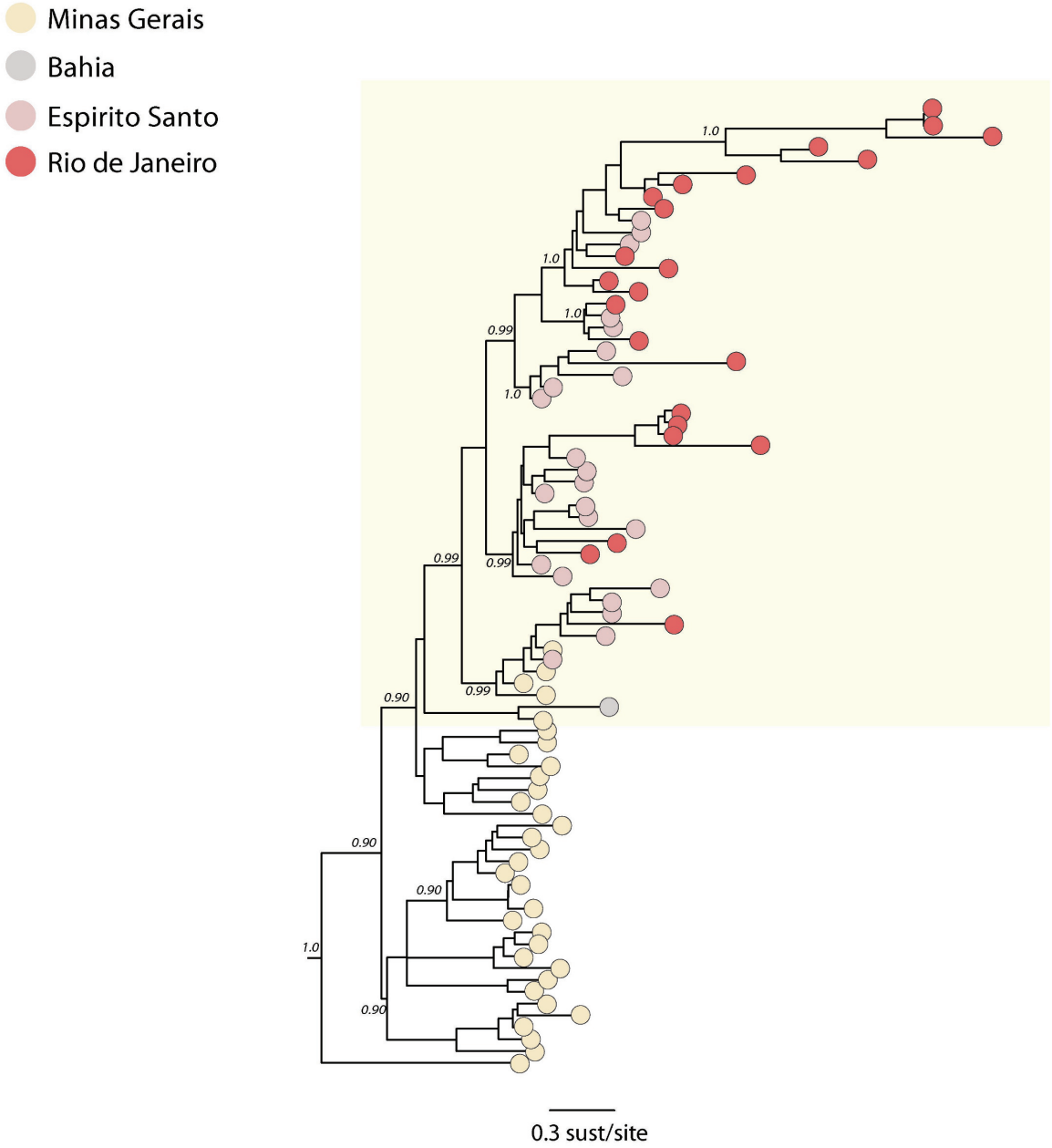
35. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, Gray RR, Arinaminpathy N, Stramer SL, Busch MP, Delwart EL. 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proceedings of the National Academy of Sciences of the United States of America 109:15066-15071.

36. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP. 2012. BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. Syst Biol 61:170–173.









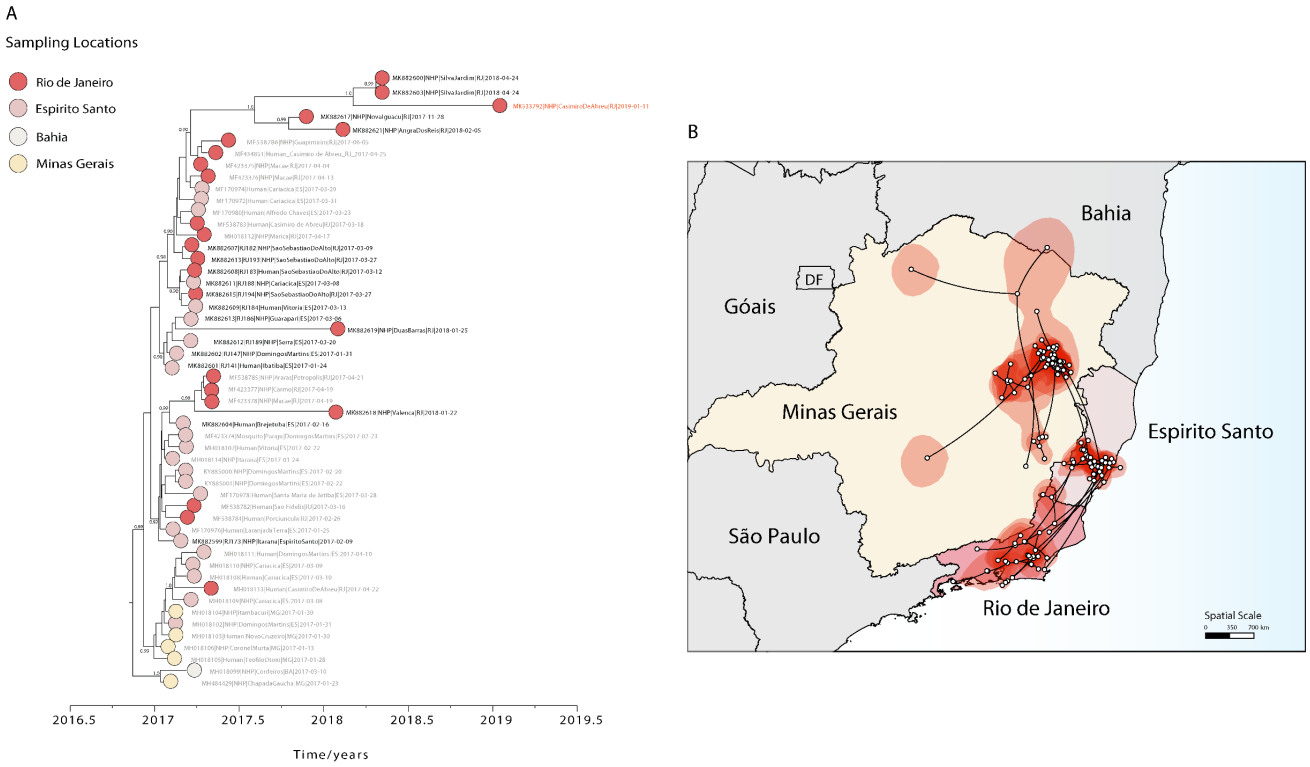


Table 1. Epidemiological data for the sequenced samples.

ID	CT value	Sample Type	Host	Species	State	Municipality	Collection Date	Age	Sex	Residence
RJ182	8,2	Liver	NHP	Alouatta Sp	RJ	São Sebastião Do Alto	09/03/2017	NA	M	-
RJ193	10,2	Liver	NHP	Alouatta Sp	RJ	São Sebastião Do Alto	27/03/2017	1	M	-
RJ141	22,4	Serum	Human	-	ES	Ibatiba	24/01/2017	16	M	Rural
RJ183	11,2	Serum	Human	-	RJ	São Sebastião Do Alto	12/03/2017	25	M	Rural
RJ194	6,5	Liver	NHP	Alouatta Sp	RJ	São Sebastião Do Alto	27/03/2017	15	F	-
RJ147	21,9	Whole blood	NHP	Alouatta Sp	ES	Domingos Martins	31/01/2017	NA	NA	-
RJ173	15	Whole blood	NHP	Cebus Sp	ES	Itarana	09/02/2017	NA	NA	-
RJ184	14,4	Liver	Human	-	ES	Cariacica	13/03/2017	65	M	Rural
RJ213	8,1	Liver	NHP	Callithrix Sp	RJ	Valença	22/01/2018	5	F	-
RJ186	10,9	Liver	NHP	Alouatta Sp	ES	Guarapari	06/03/2017	NA	NA	-
RJ177	11,5	Serum	Human	-	ES	Brejetuba	16/02/2017	46	M	Urban
RJ188	9,9	Whole blood	NHP	Callithrix Sp	ES	Cariacica	08/03/2017	NA	NA	-
RJ201	13,4	Liver	NHP	Callithrix Sp	RJ	Nova Iguaçu	28/11/2017	2	F	-
RJ219	11,2	RIM	NHP	Callithrix Sp	RJ	Angra Dos Reis	05/02/2018	NA	NA	-
RJ189	13,7	Whole blood	NHP	Alouatta Sp	ES	Serra	20/03/2017	NA	F	-
RJ216	7,2	Liver	NHP	Callithrix Sp	RJ	Duas Barras	25/01/2018	10	F	-
LABFLA09	22.1	Liver	NHP	Leontopithecus Rosalia	RJ	Silva Jardim	24/04/2018	NA	NA	-
LABFLA10	11.76	Liver	NHP	Leontopithecus Rosalia	RJ	Silva Jardim	24/04/2018	NA	NA	-

ID=study identifier; Ct=RT-qPCR quantification cycle threshold value; State= RJ-Rio de Janeiro; ES-Espirito Santo; Municipality=Municipality of residence; F=Female; M=Male; NA= Not Available.

Table 2. Sequencing statistics for the 18 new obtained sequences.

ID	Accession Number	Mapped reads	Average depth coverage	Bases covered >10x	Bases covered > 25x	Reference covered (%)
RJ182	MK882607	21104	961,34	10220	10220	99,31
RJ193	MK882613	2953	133,99	10215	9697	95,95
RJ141	MK882601	11776	523,89	10175	9955	96,17
RJ183	MK882608	1453	67,88	9934	8599	82,52
RJ194	MK882615	1146	55,27	9381	7964	79,84
RJ147	MK882602	3319	148,16	8461	7651	71,19
RJ173	MK882599	1361	63,57	9017	7628	74,68
RJ184	MK882609	1241	57,04	9480	8109	78,23
RJ213	MK882618	2520	116,01	10206	9674	93,01
RJ186	MK882610	4007	190,77	9460	9445	90,36
RJ177	MK882604	22538	1057,4	10227	10219	99,31
RJ188	MK882611	74369	3227,15	10237	10231	99,31
RJ201	MK882617	8679	399,91	9490	9454	90,34
RJ219	MK882621	8894	405,58	10205	9957	96,2
RJ189	MK882612	4840	219,19	9695	9146	89,4
RJ216	MK882619	6807	313,58	10220	9709	93,1
LABFLA09	MK882600	312871	4637,05	10210	9975	89,97
LABFLA10	MK882603	470582	5028,42	9693	9871	99,35

ID=study identifier; Accession number=NCBI accession number.