JVI Journals.ASM.org

Genomic and Phylogenetic Characterization of Brazilian Yellow Fever Virus Strains

Marcio R. T. Nunes,[a] Gustavo Palacios,[b]* Jedson F. Cardoso,[a] Livia C. Martins,[c] Edivaldo C. Sousa, Jr.,[a] Clayton P. S. de Lima,[a] Daniele B. A. Medeiros,[c] Nazir Savji,[b]* Aaloki Desai,[b] Sueli G. Rodrigues,[c] Valeria L. Carvalho,[c] W. Ian Lipkin,[b] and Pedro F. C. Vasconcelos[c,d]

Center for Technological Innovation, Instituto Evandro Chagas, Ananindeua, Brazil[a]; Columbia University, New York, New York, USA[b]; Departamento de Arbovirologia e Febres Hemorrágicas, Instituto Evandro Chagas, Ananindeua, Brazil[c]; and Universidade do Estado do Pará, Belém, Brazil[d]

Globally, yellow fever virus infects nearly 200,000 people, leading to 30,000 deaths annually. Although the virus is endemic to Latin America, only a single genome from this region has been sequenced. Here, we report 12 Brazilian yellow fever virus complete genomes, their genetic traits, phylogenetic characterization, and phylogeographic dynamics. Variable 3′ noncoding region (3′NCR) patterns and specific mutations throughout the open reading frame altered predicted secondary structures. Our findings suggest that whereas the introduction of yellow fever virus in Brazil led to genotype I-predominant dispersal throughout South and Central Americas, genotype II remained confined to Bolivia, Peru, and the western Brazilian Amazon.

Yellow fever (YF) is an infectious disease transmitted by *Culicidae* mosquitoes carrying yellow fever virus (YFV). This virus, the prototype species of the family *Flaviviridae*, genus *Flavivirus*, has a positive-sense, single-strand genome, composed of approximately 11,000 nucleotides (nt), encoding 10 viral proteins: three structural proteins (capsid [C], premembrane [PrM], and the envelope [E]) and seven nonstructural (NS) proteins (NS1-NS2A-NS2B-NS3-NS4a-NS4b-NS5-3′) (10).

YFV is endemic to tropical areas of Africa and South America, where, according to the World Health Organization (WHO), approximately 200,000 infections result in 30,000 deaths annually (31). YF is a zoonotic disease affecting nonhuman primates in the forest canopy, the primary vertebrate hosts; humans are infected upon invasion of the natural ecosystem where the virus exists. In Brazil, the virus is responsible for jungle yellow fever outbreaks with high case fatality rates. It is transmitted to humans by mosquito bites of the genera *Haemagogus* (primary) and *Sabethes* (secondary) in forested and surrounding areas (37). The two main species vectors are *Haemagogus janthinomys* and *Haemagogus leucocelaenus* (9, 37, 38).

Previous molecular epidemiology studies of YFV have provided 3′ noncoding region (3′NCR) heterogeneity analysis and phylogenetic history and viral dispersal based on partial E, NS5, and 3′NCR sequences, suggesting two genotypes in the Americas: genotype I, found in Brazil, Colombia, Ecuador, Venezuela, and the Caribbean; and genotype II, found in Bolivia, Peru, Brazil, and Ecuador (7, 8, 39). Although YFV is considered a serious public health threat in Latin America, only a single Trinidadian strain has been completely sequenced (2).

Here, we describe the full-length sequencing, genetic traits, phylogenetic analysis, and phylogeographic dynamics of 12 Brazilian YFV isolates recovered between 1980 and 2002.

MATERIALS AND METHODS

Virus strains. The YFV strains analyzed in this study are listed in Table 1 and correspond to low-passage-number isolates obtained from the World Health Organization Collaborating Center for Arbovirus Reference and Research at the Department of Arbovirology and Hemorrhagic Fevers, Instituto Evandro Chagas, Brazilian Ministry of Health.

Viral culture, nuclease treatment, and RNA extraction. Vero cells were used for virus propagation. Virus stocks were clarified by centrifugation at 3,000 rpm at 4°C and treated with nucleases to minimize contaminants.

Full-length genome sequencing. Complete genomes were obtained using high-throughput sequencing on the Genome Sequencer FLX (Roche Life Sciences, Branford, CT), as previously discussed (12, 23), alongside 5′ and 3′ rapid amplification of cDNA ends (RACE) kits (Invitrogen, Carlsbad, CA), by using specific internal primers YFV5′SP1 (5′-AKCCCTGTCTTGGGTCCAGC-3′), YFV5′SP2 (5′-TTGAACGCCTCTTGAAGGTC-3′), YFV5′SP3 (CTCTTGAAGGTCCAGGTCTA), YFV3′SP1 (5′-ATATCTGAATGGCAGCCATC-3′), and YFV3′SP2 (5′-TCACTGGCTGTTTCTTCTGCT-3′) targeting the termini. Products were cloned and sequenced in both directions with ABI Prism BigDye Terminator 1.1 cycle sequencing kits on ABI Prism 3130 DNA analyzers (Applied Biosystems, Foster City, CA).

Genome assembly. Pyrosequencing reads were trimmed to remove specific adaptors and host sequences using Newbler software (version 2.5.3; Data Processing Software Manual, 454 Life Science) and BLASTX (BLAST database and stand-alone sequence comparison software version 2.2.25), respectively. The 5′ and 3′ termini generated by Sanger sequencing were assembled to the corresponding YFV open reading frame (ORF) sequence using Geneious 5.4 (Biomatters, Auckland, New Zealand).

Genome characterization. Genomes obtained for the Brazilian YFV strains were compared with all full-length sequences publicly available, including isolates from Africa, Trinidad, and Tobago, the 17DD vaccine substrain, and vaccination-related strains (Table 1). Potential cleavage sites for polyproteins were predicted using the proteolytic processing cas-

Received 3 March 2012 **Accepted** 22 August 2012

Published ahead of print 26 September 2012

Address correspondence to Pedro F. C. Vasconcelos, pedrovasconcelos@iec.pa.gov.br, or Marcio R. T. Nunes, marcionunes@iec.pa.gov.br.

* Present address: Gustavo Palacios, United States Army Medical Institute for Infectious Diseases, Fort Detrick, Maryland, USA; Nazir Savji, New York University School of Medicine, New York, New York, USA.

M.R.T.N. and G.P., and W.I.L. and P.F.C.V., contributed equally to this article.

Supplemental material for this article may be found at http://jvi.asm.org/.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.00565-12

TABLE 1 Yellow fever virus strains isolated in Brazil, Trinidad, Africa: vaccine and vaccine-associated adverse event cases used for genetic characterization and phylogeography analyses[b]

Strain	Source of isolation	Passage history[a]	Date of isolation	Place of isolation	Coordinates	GenBank accession no.
BeAR378600	*Haemagogus* sp.	1	1980	Uruaçu, GO	−14.52, −49.15	JF912179
BeH394880	Human	1	1981	C. do Araguaia, PA	−8.25, −49.26	JF912180
BeH413820	Human	1	1983	Porto Velho, RO	−8.76, −63.9	JF912181
BeH422973	Human	1	1984	Monte Alegre, PA	−1.99, −54.08	JF912182
BeH423602	Human	1	1984	São Domingos do Capim, PA	−1.67, −47.76	JF912183
BeH463676	Human	1	1987	Breves, PA	−1.68, −50.48	JF912184
BeAR513008	*Sabethes* sp.	1	1992	Sidrolândia, MS	−20.93, −54.93	JF912185
BeH526722	Human	1	1994	Arinos, MG	−15.9, −46.1	JF912186
BeH622205	Human	1	2000	Goiás, GO	−15.92, −53.13	JF912187
BeH622493	Human	1	2000	Alto Paraíso, GO	−14.27, −47.5	JF912188
BeAR646536	*H. leucocelaenus*	1	2001	S.A. Missões, RS	−28.51, −52.23	JF912189
BeH655417	Human	1	2002	Alto Alegre, RR	2.89, −61.49	JF912190
TVP11767	*Alouatta seniculus*	–	2009	Trinidad and Tobago	10.69, −61.22	HM582851
Angola 71	Human	–	1971	Angola (Central/East Africa)	−11.2, 17.87	AY968064
Uganda 48	Human	–	1948	Uganda (Central/East Africa)	1.37, 32.29	AY968065
Couma-Ethiopia 61	Human	–	1961	Couma Ethiopia (Central/East Africa)	9.14, 40.48	DQ235229
Ivory_Coast82	Human	–	1982	Ivory Coast (West Africa)	7.53, −5.54	YFU54798
Asibi	Human	–	1927	Ghana (West Africa)	7.94, −1.02	AY640589
Gambia 01	Human	–	2001	Gambia (West Africa)	13.44, −15.31	AY572535
Ivory_Coast99	Human	–	1999	Ivory Coast (West Africa)	7.53, −5.54	AY603338
French viscerotropic	Human	–	–	–	–	YFV21056
17D	Vaccine strain	–	–	–	–	X03700
YFV-AVD2791-93	Vaccine strain	–	–	–	–	DQ118157
17-D 213	Vaccine strain	–	–	–	–	
17D-Titan	Vaccine strain	–	–	–	–	F1654700
17DD	Vaccine strain	–	–	–	–	DQ100292
French viscerotropic	Vaccine strain	–	–	–	–	YFV21056
French neurotropic	Vaccine strain	–	–	–	–	YFV 21055
YFV case 1	Vaccine adverse event	–	–	Peru	–	GQ379162
YFV case 2	Vaccine adverse event	–	–	Peru	–	GQ379163

[a] Virus passage performed in VERO cells.
[b] GO, Goias state; PA, Para state; RO, Rondonia state; RR, Roraima state; MG, Minas Gerais state; MS, Mato Grosso do Sul state; RS, Rio Grande do Sul state; Ar, arthropod; H, human; –, data not provided.

cade pattern for flavivirus ORFs developed by Rice and Strauss (33). Potential cleavage scores obtained by SignalP (http://www.cbs.dtu.dk/services/) determined host signalase cleavage sites (27). Glycosylation sites and cysteine residues were identified using NetNGlyc (version 1.0) (http://www.cbs.dtu.dk/services/NetNGlyc/) and Geneious 5.4, respectively. Highly conserved motifs and mutations were also identified.

Secondary and cyclization RNA structure prediction. Secondary and RNA cyclization structures involving the 5′NCR, the first 200 nt of the C protein, and the 3′NCR were assessed using Geneious and mfold, as previously described (45). Termini patterns were selected according to lowest fold energy and the presence of conserved secondary structures, such as 3′ and 5′ stem-looping (3′SL, 5′SL), annealing of 5′ and 3′ cyclization regions (5′ and 3′CYCs), 5′ and 3′ upstream AUG regions (5′ and 3′UAR), and the start codon at the 5′ Cap hairpin sequence (HS).

Three-dimensional protein modeling. Three-dimensional (3D) protein structures were rendered using the protein homology method by initial search and template selection using the universal protein database (templates) (http://www.rcsb.org/pdb) and applying the YFV envelope protein sequence (query sequence) as the basic parameter. The template selected (Dengue virus type 3; PDB 1UZG) was used to construct the protein model with the Swiss model server (1, 34). The predicted 3D models were run through PROCHECK (20) and ANOLEA (28) to evaluate stereochemistry parameters, ligation lengths, angle of ligation, peptidic ligation, lateral ring alignment, and rotational angle for the main and lateral chains. The final 3D structures were visualized with Visual Molecular Dynamics software (17).

Similarity analysis. Similarity among the 29 YFV genomes was calculated with the plotting similarity method in the Bootscan program in Simplot version 3.5.1 (http://sray.med.som.jhmi.edu/SCRoftware/simplot/). Phylogenetic trees were constructed using the neighbor-joining method with bootstrap resampling of 1,000 replicates using PHYLIP 3.69 (15). The Kimura 2-parameter model was used as a nucleotide substitution model. The phylogenetic permutation values (ppt) expressed on the *y* axis were calculated from bootstrap values by using a 50% threshold. The *x* axis represents the YFV genome positions. Bootscanning was conducted in a slide window of 200 nt with 20-nt steps. The Trinidadian strain TVP 11767 (HM582851) was used as the reference (2).

Phylogenetic and phylogeography analyses. Phylogenetic analysis was conducted for 20 complete YFV genomes retrieved from GenBank. From these analyses, vaccine and vaccine-related strains were removed from the data set to avoid bias of geographic position. Before phylogenetic analysis, possible recombination events were identified using both a genetic algorithm for recombination detection (GARD) (19) and the Phi test (6), available in SplitsTree version 4 (18). Sequences were aligned using the muscle algorithm implemented in Geneious 5.4. Maximum likelihood (ML) phylogenies were constructed with the MrBayes-Multi version 3.1.2 (32), applying the general time reversible nucleotide substitution model (GTR + 4Γ) and optimizing across site rate variation. The

TABLE 2 Genome size of YFV Brazilian strains including the 5′NCR, ORF, and 3′NCR

Strain[a]	Size (nt)			
	5′NCR	ORF	3′NCR	Total
BeAr 378600	118	10.236	535	10.889
BeH 394880	118	10.236	654	11.008
BeH 413820	118	10.236	441	10.795
BeH 422973	118	10.236	506	10.860
BeH 423602	118	10.236	479	10.852
BeH 463676	118	10.236	650	11.004
BeAr 513008	118	10.236	654	11.008
BeH 526722	118	10.236	650	11.004
BeH 622205	118	10.236	654	11.008
BeH 622493	118	10.236	463	10.817
BeAR 646536	118	10.236	654	11.008
BeH 655417	118	10.236	505	10.859

[a] Ar, arthropod; H, human.

parameters of a full probabilistic model, including timed sequence evolution and spatial-temporal dispersal, were estimated using a Bayesian statistical inference approach implemented in BEAST (13, 14, 21). The Bayesian skyline coalescent prior (36) was used to model fluctuations of the effective population size over time. Bayesian phylogenetic inference was also applied to GTR 4 DNA substitution model to account for among-site rate variation. Isolation dates in years were available from GenBank and laboratory records. The spatial-temporal dispersal of YFV between Africa and the Americas was reconstructed using the discrete model and within the Americas using the continuous probabilistic model of viral diffusion, both implemented in the BEAST package. For each alignment, three Markov chain Monte Carlo (MCMC) analyses were run for 50 million cycles, sampling every 10,000th state until adequate mixing was achieved. Bayesian phylogenetic analyses were computed using the BEAGLE library (5) to augment the computational speed. After removing 10% of the burn-in, runs were combined with LogCombiner (http://tree.bio.ed.ac.uk/software). Maximum clade credibility (MCC) trees were summarized with Tree Annotator and visualized using Fig Tree. The spatial phylogenetic reconstruction of evolutionary dynamics (SPREAD) application (5) was used to visualize and convert the estimated divergence times and spatial estimates annotated in the MCC trees to a keyhole markup language file (KML) for visualization in the Google Earth virtual software (www.google.com/earth). All evolutionary parameters are reported as posterior means, along with the correspondent 95% Bayesian credible intervals (95% BCI).

RESULTS

Genome characterization. The Brazilian YFV genomes ranged in size from 10,795 to 11,008 nt. The predicted ORFs were 10,236 nt (3,411 amino acids [aa]) flanked by a conserved 5′NCR consisting of 118 nt and highly variable 3′ termini ranging from 441 to 654 nt in length (Table 2). Genetic divergence ranged from 2% for 5′NCRs to 12% for the M gene and 2K signal peptide, with 10% divergence across the entire genome. Among South American strains, 4.8% of genetic variability was observed, while between South American and African strains it was 16% (see Table S1 in the supplemental material). By the genetic similarity plot method, four distinct groups were depicted (permutation values over 95% within groups and >75% between groups). Groups I to IV are shown in Fig. S1 in the supplemental material, and unique mutations for YFV South American strains are shown in Fig. S2 in the supplemental material.

Potential cleavage sites, cysteine residues, and glycosylation sites. Potential cleavage sites were predicted for each gene junction across the ORF of each YFV strain. Highly conserved motifs representing potential sites for protease cleavage activity were identified (see Table S2 in the supplemental material). The numbers of cysteine residues present throughout the viral proteins were as follows: AncC, 0; M, 6; E, 12; NS1, 12; NS2A, 3; NS2B, 0; NS3, 10; NS4A/2k, 1; NS4B, 3; NS5, 17. Glycosylation sites were found in three viral proteins: M (residues M134, M150, and M266), E (residues E554, E594), and NS1 (NS1-908, NS1-986). The number and position of cysteine residues and glycosylation sites were conserved among all YFV strains.

Analysis of the 5′ and 3′ ends. Alignment of the 5′ and 3′ ends revealed 5′NCR conservation and 3′NCR variability.

In the 5′NCR, minor nucleotide mutations were observed in comparison to other YFV strains [A80G, C108T, (T/C)118A]. Unique to BeH 413820 and BeH 422973 are nucleotide substitutions at position 117 (A→C) and 109 (T→A), respectively.

In the 3′NCR, size heterogeneity was observed as previously described (8). Long deletions were found in the Brazilian YFV strains in comparison to African and vaccine/vaccine-derived strains herein denominated South American deleted motif YFVSADM1. These deletions were observed among the YFV repetition motifs RYF1 and RYF2. Furthermore, specific sequences for Brazilian YFV strains, namely, YFV conserved motifs YFVSACM1 and YFVSACM2, were also found. Three YFV strains (Be H6554117, BeH 622493, and BeH 422973) contained a unique motif (YFVSAUM). The Brazilian strains BeH655417, BeH622493, BeH422973, and BeH423602 do not have the YFVSADM2 motif in domain III within the 3′SL (Table 3).

The absence of YFV repetition motifs RYF1 and RYF2 (YFVSADM1) and the presence of RYF3, conserved sequence 2 (CS2), 3′UAR, and 3′CYC sequences was observed (8, 24). Furthermore, the pentanucleotides 3′CACAG observed in the 3′SL of flaviviruses were absent from strains with the YFVSADM2 deletion.

Except for strain BeH 413820, the 3′CYC imperfect terminal (3′CYC imp) sequence was identified in all Brazilian strains. Five strains (BeH 655417, BeH 622493, BeH 422973, BeH 423602, and BeAR 378600) had the 3′CYC motif deleted within the YFVSACM1 motif. In general, a single nucleotide change (A→G)

TABLE 3 3′ Noncoding region motifs observed in South American YFV strains

3′NCR motif	Sequence (5′→3′)
YFVSADM1	GGGATACAAACCACGGGTGGAGAACCGGACTCCCCACAACCTGAAACCGGGATAT AAACCACGGCTG
YFVSACM1	TGTCGGCCCAGAAACCCGCGTGGGTTCTGCCACTGCTAAGCTGTGAGGCAGTGCAGGCTGGGA TAGCCGCCTTCC ATGTTGCGAAAAACATGGTTTCTGAGACCTCCC
YFVSACM2	A[G]CCTCAGAGTAAAACCAAATGGAGCCTCCGCTACCACCTCCCCACGTGGTGGTAGAAAGACGGGGTCTAGAGGTTA GAGGAGACCCTCCAGGGAACAAATAGTGGGACCATATTG
YFVSAUM	ACCTCAGAGTAAAACCAAATGGAGCCTCCGCTACCACCTCCCCACGTGG

(a)

Strains	Genotype	Cyclization Pattern
BEH413820	II	II
BEH655417	I	III
BEH622493	I	III
BEH422973	I	III
BEH423602	I	IV
BEAR378600	I	V
BEH622205	I	VI
BEAR646536	I	VI
BEH526722	I	VI
BEH463676	I	VI
BEH394880	I	VI
BEAR513008	I	VI

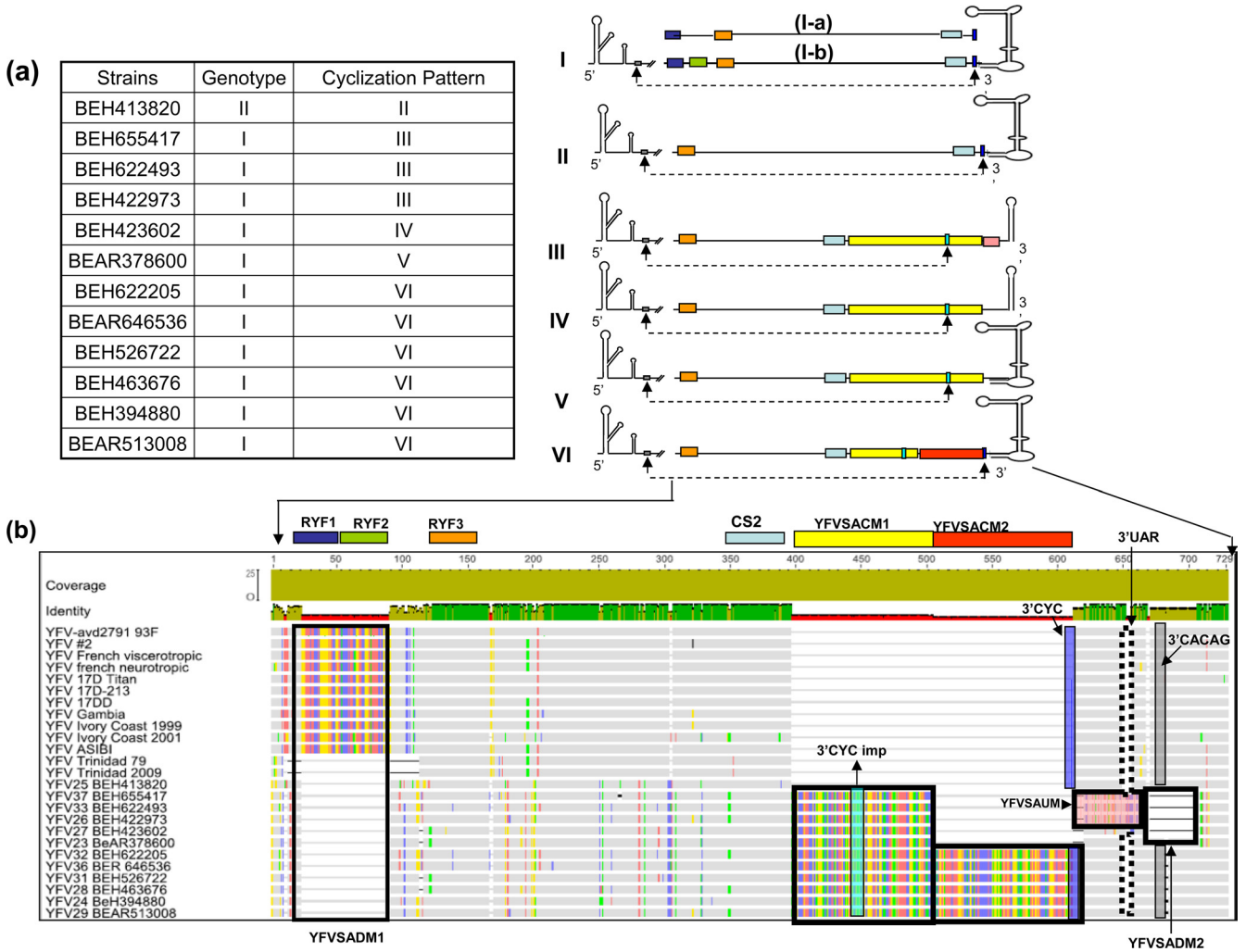

FIG 1 (a) Brazilian YFV strains according to the genotype and cyclization patterns; (b) schematic representation of the five distinct patterns found for 3′NCR of 12 YFV Brazilian isolates. (II) 3′NCR/deleted YFVSACM1 and YFVSACM2; (III) long 3′NCR Uniq motif-plus (YFVSAUM)/deleted YFVSADM2; (IV) long 3′NCR/deleted YFVSADM2; (V) long 3′NCR/YFVSADM1; (VI) extralong 3′NCR (YFVSACM1/YFVSACM2). Pattern I corresponds to African strains exhibiting the RYF1, RYF2, RYF3, CS2, and 3′CYC motifs (I-a; West African strains) or lack of RYF2 repetition sequence (I-b; East and Central African strains). (c) Predicted secondary structures generated by the mfold program for the five distinct patterns (II to VI) described for YFV strains isolated in Brazil. Conserved structures are indicated by arrows, brackets, colors, or boxes. Abbreviations: CS, conserved sequences; RCS, repeated conserved sequence; SmLs, small loop; LSH, long stable hairpin; 3′CYC, cycling sequence within CS1 of the 3′NCR; 5′CYC, cycling sequence within the capsid (Cap) gene; HS, hairpin sequence; UAR, upstream AUG region; NCR, noncoding region.

was observed; however, for strain BeH 394880, another mutation (T→G) was also observed.

In addition, five distinct 3′NCR patterns (II to VI) were predicted in Brazilian YFVs: 3′NCR/deleted YFVSACM1 and YFVSACM2, represented by genotype II (pattern II); long 3′NCR Uniq motif-plus(YFVSAUM)/deleted YFVSADM2 (pattern III); long 3′NCR/deleted YFVSADM2 (pattern IV); long 3′NCR/YFVSADM1 (pattern V); and extralong 3′NCR (YFVSADM1/YFVSADM2) (pattern VI). Pattern I correspond to the African strains exhibiting the RYF1, RYF2, RYF3, CS2, and 3′CYC motifs (Fig. 1 and 2).

In all patterns, the 5′ secondary structures, represented by a long SL with a lateral loop, secondary hairpin, and start codon hairpin, were conserved. The 3′ secondary structure was not conserved in strains represented by patterns III and IV (Fig. 1).

3D structure prediction of yellow fever virus envelope protein. The analysis of the 12 YFV E protein sequences revealed point mutations compared to both ASIBI and 17D/Titan vaccine strains across the three E protein domains (DI, DII, and DIII).

In DI, amino acid changes were observed at positions E62 (N→S), E67 (H→N; except for strain BeH 413820), E83 (A→E, except for strain BeH 413820), E177 (K→R for strains BeH 422973, BeAR 513008, BeH 622205, BeH 622493, BeAR 646536, and BeH 655417), E191 (G→S), E198 (M→R for strain BeH 463676), E207 (R→K, except for BeH 413820), E243 (R→K), E268 (T→A for strain BeH 413820), tripeptide 270-272 (DNN→DSK or GSK for strain BeH 413820), and E282 (S→A for strains BeH 394880 and BeH 413820).

For DII, amino acid substitutions were found at E120 (T→A for strain BeH 422973), E147 (K→V for strain BeAR 646536), and E154 (T→A for strain BeH 413820).

DIII of the YFV strains revealed mutations at positions E318

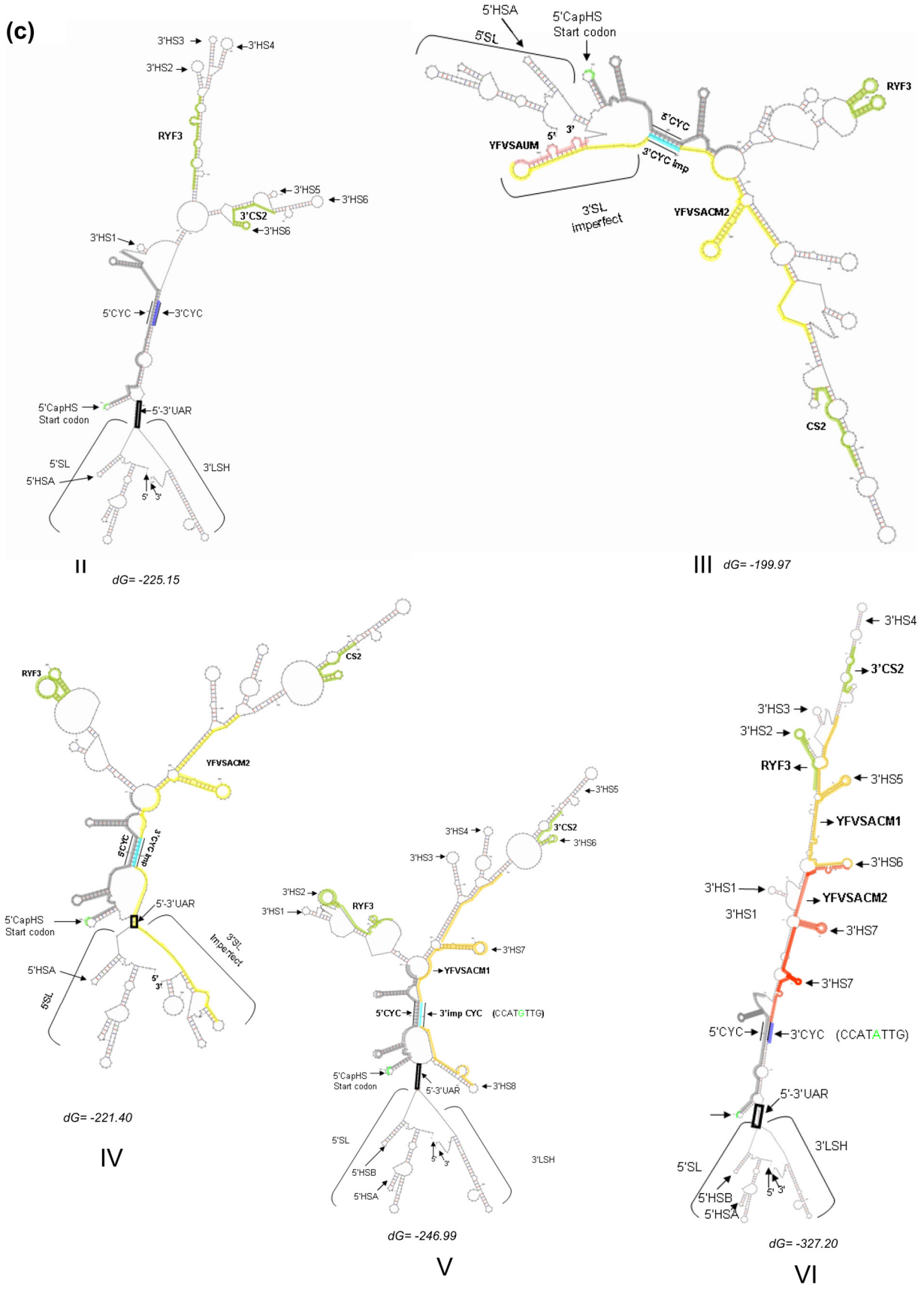

FIG 1 continued

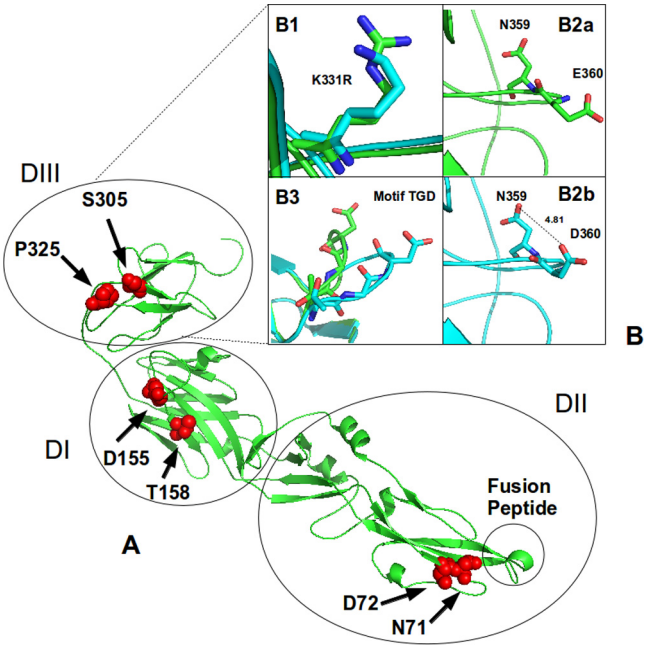

FIG 2 (A) Predicted 3D structure of the YFV E protein. The protein domains I, II, and III and fusion-peptide subdomain are shown in circles. Neutralizing epitopes are represented as red spheres along the E protein domains. (B) The E protein DIII is highlighted; comparison between ASIBI (blue) and YFV BeH 413820 (green) 3D structures. (B1) Conformational changes in YFV BeH 413820 strain due to amino acid substitution (K331R). (B2a) YFV BeH 413820 strain; N359 and E360 amino acids related to rescue of virulence in attenuated strains. (B2b) Amino acid changes in the DIII virulence-related epitope E360D. Note the hydrogen bond (dotted line) between amino acids N359 and D360 in the YFV strain BeH 413820 that are related to virulence. The number 4.81 corresponds to the distance in angstroms revealing the presence of a hydrogen bond, which is associated with viral persistence. (B3) TGD motif revealing conformational changes for ASIBI (blue) and the wild-type YFV BeH 413820 strain (green).

(V→A), E331 (K→R), E335 (I→M), E344 (I→V), and E360 (D→E for strain BeH 413820).

The DII fusion peptide (DRGWGNGCGLFGK), as well the DIII TGD motifs, was conserved in all YFV strains. Mutations across the E protein domains and 3D structure are presented in the Fig. S2 in the supplemental material.

Phylogeographic analysis. The Bayesian MCC complete genome analysis demonstrated the presence of two major groups (I and II), where group I represented East/Central African strains, and II included Western Africa, Trinidadian, and Brazilian strains. The time of emergence for the most recent common ancestor (TMRCA) for the South American YFV strains was estimated to be 332 years ago (95% Bayesian credibility interval, 210 to 519) from a Western African lineage ancestor. The predicted substitution rates for the genomes were 1.12E−3 substitutions/site/year (East African strains), 3.1E−4 substitutions/site/year (West African strains), 3.04 E−4 substitutions/site/year (Brazilian strains), and 9.3 E−4 substitutions/site/year (Trinidadian strain) (Fig. 3a).

The phylogeographic model demonstrated the introduction of YFV from Africa to the Americas (Fig. 3b), and the continuous model indicates that YFV dispersion in South America is relatively recent. The virus was likely introduced in Northern/Northeast Brazil before dispersing to other Brazilian regions (Western/Central and Southern/Southeast regions) and Trinidad and Tobago at a dispersion speed of 24.95 km/year (95% Bayesian credibility interval (13.6 to 37.4) (Fig. 3a, b, and c).

DISCUSSION

Despite Brazil being the largest country where YF is endemic, with disease emerging in local regions that were previously considered YF-free for almost 50 years (8, 9, 38), no Brazilian YFV complete genomes have been reported (2, 3, 26, 30, 42, 43). Previous genomic, phylogenetic, and evolutionary analyses were based on partial nucleotide sequences restricted to the prM/M E and NS5/3′NCR junctions (7, 8, 11, 22, 39).

Although sequence analysis of YFV isolates collected from various hosts in distinct geographic locations at different times have yielded insights into 3′NCR heterogeneity and evolutionary history (8, 39), full genome sequencing reported here has the potential to enhance diagnostics and provide a more comprehensive view of the phylodynamics of YFV.

The 12 Brazilian wild-type YFV strains represented here showed high levels of genetic divergence and size heterogeneity (ranging from 10,795 nt to 11,008 nt in length) (Table 2) as previously described (8, 29).

Gene analysis revealed conserved motifs among all strains. Interestingly, many genetic signatures were found in Brazilian YFV strains compared to in African, Trinidadian, and vaccine-associated strains (see Fig. S2 in the supplemental material). These unique motifs should be useful for development of molecular probes and PCR-derived methods (e.g., real-time PCR, mass tag-PCR) capable of differentiating wild-type YFV cases from suspected severe adverse events following vaccination mainly when epidemiological history is poorly understood (e.g., lack of vaccine shot information, period of vaccination, response to vaccine). Potential cleavage sites for generation of YFV proteins, N-glycosylation sites, and cysteines were conserved in all Brazilian YFV genomes, suggesting their functional roles are conserved (4).

Genetic variability among the Brazilian YFV strains was observed along the entire genome (4.8%,); however, greater divergence was observed in the 3′NCR region (6%) as previously described, suggesting a large amount of size heterogeneity in these regions (8, 39).

Five distinct 3′NCR patterns were described (Fig. 1), three of them (III, IV, and V) presented a long conserved motif (YFVSAC M1), and two (III and IV) revealed a 37-nt deletion corresponding to the YFVSADM2 motif. Pattern VI presented an extralong YFV SACM2 motif. Despite the lack of 3′CYC sequences in YFV strains belonging to the 3′NCR patterns III, IV, and V, the RNA cyclization between 5′ and 3′NCR termini were made using the 3′CYCimp. In the case of pattern IV strains where the 3′CYC regular motif is preserved, the imperfect 3′ cyclization sequence was not associated with secondary structure.

The 37-nt deletion (YFVSADM2) appears to modify the secondary structure of the 3′SL in patterns III and IV, including 3′CACAG pentanucleotide. All flavivirus genomes contain a 3′SL secondary structure constituted by the most downstream 100 nt of the viral RNA genome and 3′CACAG pentanucleotide at the top of the 3′SL structure. These structures are required for virus replication and binding to both virus-coded and cellular proteins (44), and modifications in the 3′SL or 3′CACAG pentanucleotides

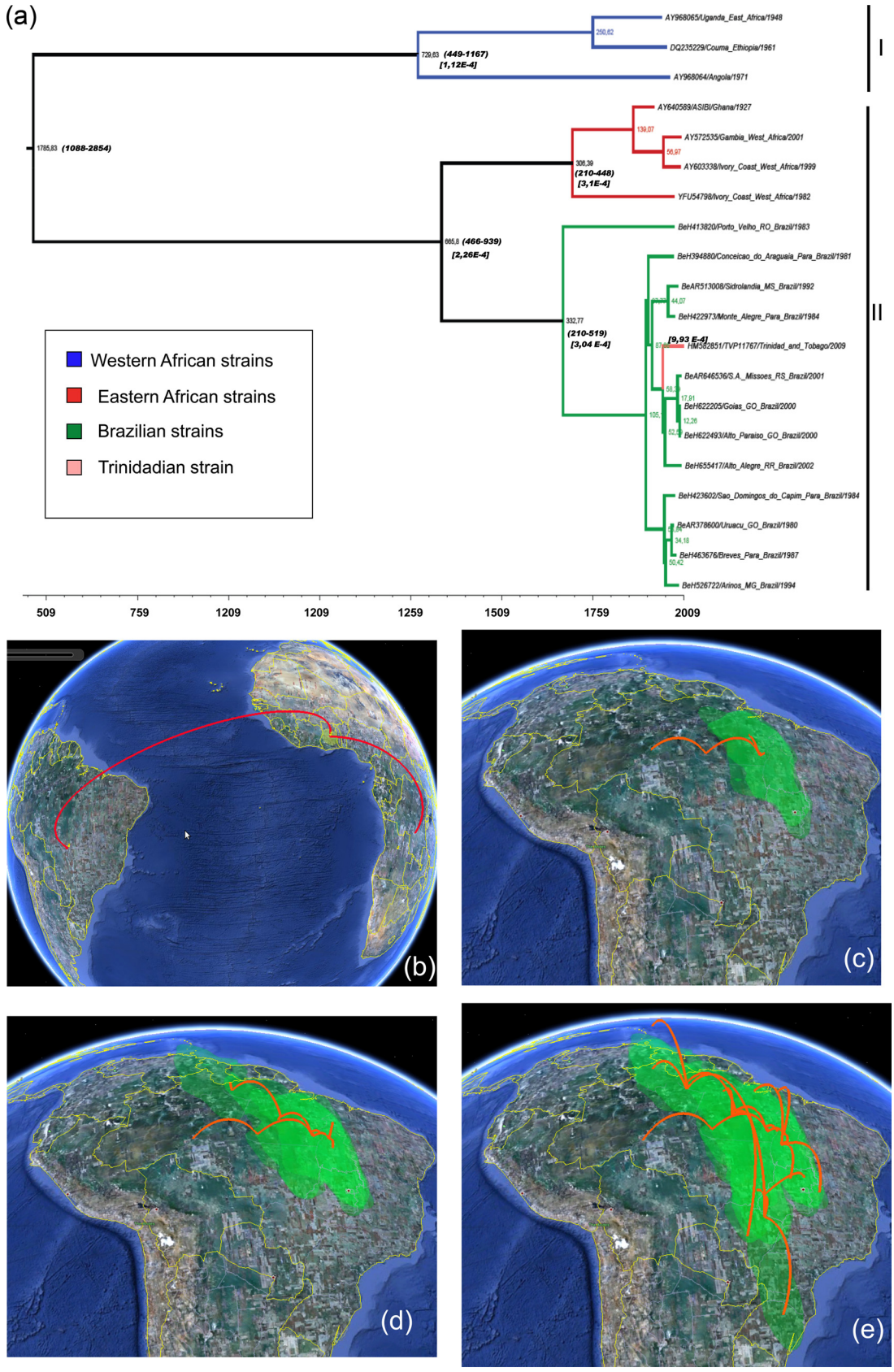

FIG 3 (a) Bayesian maximum clade credibility tree demonstrating the phylogenetic relationships among the YFV complete genomes. Major groups are indicated (I and II). Subgroups according to geographic location (West Africa, East Africa, Trinidad, and Brazil) are colored. Numbers under the bar represent the timing scaled for the most recent common ancestor. Node ages and substitution rates (substitutions/site/year) are placed over each main node in the tree. Numbers within parentheses and brackets represent the 95% HPD Bayesian posterior probabilities and dispersion rates expressed in percentages, respectively. Spatiotemporal diffusion of YFV throughout Africa to the Americas (b). Temporally framed snapshots of the dispersal patterns and respective 95% Bayesian credibility regions (BCR) are indicated (light-green areas) for the years 1940 (c), 1970 (d), and 2000 (e) (Maps in panels b to e reproduced from Google Earth [www.earth.google.com/].)

may result in viral attenuation (35). Further studies are needed to investigate all 3′ motifs to determine whether the YFVSADM2 deletion affects replication competency.

Mutations observed in the amino acid positions N62S, G191S, R243K, tripeptide 270-272, S282A, V318A, I335M, compared to the Asibi strain are genetic signatures of strains circulating in the Americas (41). Although the H67N, A83E, T154A, K177R, and R207K mutations are not considered genetic signatures, they are commonly found in American strains. Furthermore, the K331R mutation was observed in the Brazilian YFV strains and is suggested to be associated with viscerotropism in hamsters (24, 25), while the D360E substitution is associated with virulence rescue in attenuated strains (40). The role of other mutations (T120A, A147V, M198R, and T268A) remains uncharacterized.

Genetic analyses of variability among complete YFV genomes and phylogenetic analysis support previous observations regarding the existence of distinct genetic groups in Africa and in the Americas (8, 11, 29, 39, 42) and the existence of at least two distinct genetic lineages or genotypes (I and II) cocirculating in Brazil. The complete genome phylogenetic analysis indicates that the Brazilian strains are closely related to Western African strains, suggesting an African origin (Fig. 3a) approximately 332 years ago (95% highest posterior density [HPD], 210 to 519) in the year of 1677. This time is compatible with previous publications based on E gene analyses and also is in accordance with the first report of a yellow fever outbreak in Brazil in 1685, when the virus was possibly introduced in Recife, the capital of Pernambuco state, Northeast Brazil, during slave transportation from Sao Tome, Africa, to Brazil, with one stop in Santo Domingo, Antilleans, where the disease was devastating the population (16).

The phylogeographic models (discrete and continuous) confirmed the introduction of the YFV in Brazil from West Africa and suggest that American dispersion is relatively recent (Fig. 3a). Previous geographical dispersion analysis was performed using partial sequences (2), while the current results, using complete genomes, suggest that YFV entered the Americas in Northern/Northeast Brazil before dispersing to other regions in Brazil and Trinidad at 24.95 km/year (95% BCI of 13.6 to 37.47) with a mutation rate of 3.04E−4 (1.924E−4, 4.24E−4) (Fig. 3c, d, e), similar to West African strains and approximately 10 times lesser than Eastern African strains ($1.12 \times 10E3$). It likely became endemic in areas where the virus could be maintained among vertebrate and invertebrate reservoirs, such as the Brazilian Amazon and central-western region of Brazil (9). These rates (dispersion and mutation rates) are within the confidence interval for rates estimated for the envelope and NS5 genes of YFV and other flaviviruses (8) and correspond to the first estimations using entire genomes of YFV.

In conclusion, this study supports the existence of previously reported genotypes circulating in Brazil (genotypes I and II) and provides informative data regarding the virus's African origin and dispersion throughout the Americas. Further studies are needed to complete genomes from other countries where YFV is endemic to elucidate the dynamics of geographic dispersion of YFV throughout the world. The availability of the complete genomic YFV sequence may facilitate further phenotypic evaluation of viral differences, particularly the potential significance of variable regions in the C, NS5, and 3′NCR by using infectious clones. It may also enable improvements in YFV diagnostics.

ACKNOWLEDGMENTS

We thank Keley Nascimento Barbosa Nunes (Center for Technological Innovation) for the technical assistance with the GS FLX 454 and Ana Cecilia Ribeiro Cruz (Department of Arbovirus and Hemorrhagic Fevers) for sequencing part of the envelope gene of the Brazilian YFV strains. We also thank Nuno Rodrigues Faria (Department of Microbiology and Immunology, Katholieke Universiteit Leuven, Leuven, Belgium) for his helpful discussions about phylogeography.

We declare no conflicts of interests.

This study was supported by CNPq grants (302987/2008-8 and 301641/2010-2), CNPq/CAPES/FAPESPA (573739/2008-0), and awards from the National Institutes of Health (AI57158, NBC-Lipkin), USAID Predict, and the Defense Threat Reduction Agency.

REFERENCES

1. **Arnold K, Bordoli L, Kopp J, Schwede T.** 2006. The SWISS-MODEL workspace: a Web-based environment for protein structure homology modelling. Bioinformatics **22**:195–201.
2. **Auguste AJ, et al.** 2010. Yellow fever virus maintenance in Trinidad and its dispersal throughout the Americas. J. Virol. **84**:9967–9977.
3. **Bae HG, et al.** 2005. Analysis of two imported cases of yellow fever infection from Ivory Coast and The Gambia to Germany and Belgium. J. Clin. Virol. **33**:274–280.
4. **Beasley DW, et al.** 2005. Envelope protein glycosylation status influences mouse neuroinvasion phenotype of genetic lineage 1 West Nile virus strains. J. Virol. **79**:8339–8847.
5. **Bielejec F, Rambaut A, Suchard MA, Lemey P.** 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics **15**: 2910–2912.
6. **Bruen TC, Philippe H, Bryant D.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172**:2665–2681.
7. **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**(5):e75. doi:10.1371/journal.ppat.0030075.
8. **Bryant JE, et al.** 2005. Size heterogeneity in the 3′ noncoding region of South American isolates of yellow fever virus. J. Virol. **79**:3807–3821.
9. **Cardoso JC, et al.** 2010. Yellow fever virus in Haemagogus leucocelaenus and Aedes serratus mosquitoes, southern Brazil, 2008. Emerg. Infect. Dis. **16**:1918–1924.
10. **Chambers TJ, Hahn CS, Galler R, Rice CM.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. **44**:649–688.
11. **Chang G-JJ, Cropp CB, 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**:5773–5780.
12. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:394–403.
13. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG.** 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. **22**:1185–1192.
14. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biol. **7**:214.
15. **Felsenstein J.** 1989. PHYLIP—Phylogeny Interference Package (version 3.2). Cladistics **5**:164–166.
16. **Franco O.** 1969. History of the yellow fever in Brazil. Rev. Bras. Malariol. Doencas Trop. 21:315–512. [Article in Portuguese.]
17. **Humphrey W, Dalke A, Schulten K.** 1996. VMD: visual molecular dynamics. J. Mol. Graphics. **14**:33–38.
18. **Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254–267.
19. **Kosakovsky W, et al.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **10**:1891–1901.
20. **Laskowiski R, Macarthur M, Moss D, Thornton J.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallog. **26**:283–291.
21. **Lemey P, Rambaut A, Drummond AJ, Suchard MA.** 2009. Bayesian phylogeography finds its roots. PLoS Computat. Biol. **5**(9):e1000520. doi: 10.1371/journal.pcbi.1000520.
22. **Lepiniec L, et al.** 1994. Geographic distribution and evolution of yellow

fever viruses based on direct sequencing of genomic cDNA fragments. J. Gen. Virol. **75**:417–423.
23. **Margulies M, et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380.
24. **McArthur MA, Suderman MT, Mutebi Xiao J-PSY, Barrett AD.** 2003. Molecular characterization of a hamster viscerotropic strain of yellow fever virus. J. Virol. **77**:1462–1468.
25. **McArthur MA, Xiao SY, Barrett AD.** 2005. Phenotypic and molecular characterization of a non-lethal, hamster-viscerotropic strain of yellow fever virus. Virus Res. **110**:65–71.
26. **McElroy KL, Tsetsarkin KA, Vanlandingham DL, Higgs S.** 2005. Characterization of an infectious clone of the wild-type yellow fever virus Asibi strain that is able to infect and disseminate in mosquitoes. J. Gen. Virol. **86**:1747–1751.
27. **Medeiros DB, Nunes MR, Vasconcelos PF, Chang GJ, Kuno G.** 2007. Complete genome characterization of Rocio virus (Flavivirus: Flaviviridae), a Brazilian flavivirus isolated from a fatal case of encephalitis during an epidemic in Sao Paulo state. J. Gen. Virol. **88**:2237–2246.
28. **Melo F, Feytmans E.** 1998. Assessing protein structures with a non-local atomic interaction energy. J. Mol. Biol. **277**:1141–1152.
29. **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.
30. **Pisano MR, Mercier V, Deubel V, Tolou H.** 1999. Complete nucleotide sequence and phylogeny of an American strain of yellow fever virus, TRINID79A. Arch. Virol. **9**:1837–1843.
31. **Robertson SE, et al.** 2006. Yellow fever: a decade of reemergence. JAMA **14**:1157–1162.
32. **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572–1574.
33. **Rice C, Strauss J.** 1990. Production of flavivirus polypeptides by proteolytic processing. Semin. Virol. **1**:357–367.
34. **Schwede T, Kopp J, Guex N, Peitsch MC.** 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. **31**: 3381–3385.
35. **Silva PA, et al.** 2007. Conservation of the pentanucleotide motif at the top of the yellow fever virus 17D 3′ stem-loop structure is not required for replication. J. Gen. Virol. **88**:1738–1747.
36. **Suchard MA, Rambaut A.** 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics **25**:1370–1376.
37. **Vasconcelos PF.** 2003. Yellow fever. Rev. Soc. Bras. Med. Trop. **36**:275–293. [Article in Portuguese.]
38. **Vasconcelos PF.** 2010. Yellow fever in Brazil: thoughts and hypotheses on the emergence in previously free areas. Rev. Saude Publ. **44**:1144–1149.
39. **Vasconcelos PF, et al.** 2004. Genetic divergence and dispersal of yellow fever virus, Brazil. Emerg. Infect. Dis. **10**:1578–1584.
40. **Vlaycheva LA, Chambers TJ.** 2002. Neuroblastoma cell-adapted yellow fever 17D virus: characterization of a viral variant associated with persistent infection and decreased virus spread. J. Virol. **76**:6172–6184.
41. **Wang E, et al.** 1995. Comparison of the genomes of the wild-type French viscerotropic strain of yellow fever virus with its vaccine derivative French neurotropic vaccine. J. Gen. Virol. **76**:2749–2755.
42. **Wang H, et al.** 1997. Genetic variation among strains of wild-type yellow fever virus from Senegal. J. Gen. Virol. **78**:1349–1355.
43. **Whittembury A, et al.** 2009. Viscerotropic disease following yellow fever vaccination in Peru. Vaccine **27**:5974–5981.
44. **Yu L, Markoff L.** 2005. The topology of bulges in the long stem of the flavivirus 3′ stem-loop is a major determinant of RNA replication competence. J. Virol. **79**:2309–2324.
45. **Zuker M, Mathews D, Turner D.** 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, p 11–43. *In* Barciszewski J, Clark B (ed), RNA biochemistry and biotechnology. Kluwer Academic Publishers, Amsterdam, Netherlands.