Am. J. Trop. Med. Hyg., 91(2), 2014, pp. 225–234
doi:10.4269/ajtmh.13-0523

Phylogeography and Molecular Epidemiology of an Epidemic Strain of Dengue Virus Type 1 in Sri Lanka

Karen E. Ocwieja, Anira N. Fernando, Scott Sherrill-Mix, Sesh A. Sundararaman, Rashika N. Tennekoon, Rashmi Tippalagama, Shivankari Krishnananthasivam, Gayani Premawansa, Sunil Premawansa, and Aruna Dharshan De Silva*

University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; Genetech Research Institute, Colombo, Sri Lanka; North Colombo Teaching Hospital, Ragama, Sri Lanka; Department of Zoology, University of Colombo, Colombo, Sri Lanka; Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, California

Abstract. In 2009, a severe epidemic of dengue disease occurred in Sri Lanka, with higher mortality and morbidity than any previously recorded epidemic in the country. It corresponded to a shift to dengue virus 1 as the major disease-causing serotype in Sri Lanka. Dengue disease reached epidemic levels in the next 3 years. We report phylogenetic evidence that the 2009 epidemic DENV-1 strain continued to circulate within the population and caused severe disease in the epidemic of 2012. Bayesian phylogeographic analyses suggest that the 2009 Sri Lankan epidemic DENV-1 strain may have traveled directly or indirectly from Thailand through China to Sri Lanka, and after spreading within the Sri Lankan population, it traveled to Pakistan and Singapore. Our findings delineate the dissemination route of a virulent DENV-1 strain in Asia. Understanding such routes will be of particular importance to global control efforts.

INTRODUCTION

Dengue virus (DENV) is a mosquito-borne flavivirus found in tropical and subtropical regions. The main mosquito vector, *Aedes aegypti*, thrives in urban areas across these regions, although rural areas are increasingly affected.[1] The reported incidence of DENV infection has increased 30-fold since the 1960s, and the World Health Organization (WHO) estimates that over 50 million DENV infections occur annually.[1] A recent study suggested that this number may be as high as 390 million per year.[2] Each year, approximately 500,000 patients, the majority being children, are hospitalized for dengue hemorrhagic fever (DHF), with a mortality rate of up to 5%.[1,3]

DENVs fall into four genetically divergent serotypes (DENV-1, -2, -3, and -4). Infection by any one of these viruses confers lifelong immunity to viruses of that serotype but only transient cross-protective immunity to viruses of other serotypes.[4] DENV infection may be asymptomatic or symptomatic, with presentations ranging from mild febrile illness with myalgias to severe life-threatening disease (DHF or dengue shock syndrome [DSS]).[4–6] The latter presentation predominantly occurs on secondary infection with a DENV serotype to which the individual had not been previously exposed. Partly as a result of increased travel, viruses of all four serotypes now circulate in much of the tropical and subtropical world, setting the stage for increased incidence of severe disease.[2,7,8]

In Sri Lanka, viruses of serotypes 2 and 3 predominated for much of the last decade, with DENV-3 causing epidemics in 2002 and 2004.[9,10] In 2009, the arrival of a new DENV-1 viral strain coincided with the largest DHF epidemic (346 deaths) since dengue was made a reportable disease in 1996.[11,12] Dengue disease has reached epidemic proportions in Sri Lanka every year since 2009,[13] and DENV-1 has remained the predominant disease-causing serotype (Sirisena[14] and A. D. De Silva, unpublished data). In 2012 in Sri Lanka, there were 44,456 reported cases of clinically diagnosed dengue disease, of which over 10,000 cases were in the capital city of Colombo.[13] It is unclear where the 2009 epidemic DENV-1 came from, whether the same strain has persisted in Sri Lanka, or why it caused so much severe disease.

Given the unavailability of specific treatment or vaccine for DENV, it is paramount that we build our understanding of the dissemination routes of the virus between endemic areas. Such data could predict at-risk areas during future pandemics, inform mosquito control efforts, and direct delivery of drugs and vaccines should they become available. In this work, we use full-genome sequencing and phylogenetic methods to characterize the spread of a particularly virulent strain of DENV-1 within the Asian subcontinent and investigate its continued circulation in Sri Lanka.

MATERIALS AND METHODS

Study population and sample collection. The study was approved by the Ethics Review Committee, Faculty of Medicine, University of Colombo and the Institutional Review Board of the University of Pennsylvania (Institutional Review Board 7 project 816053). All patients were provided with detailed information about the study, and written consent was obtained before the patient was recruited for the study. All patients recruited were above the age of 18 years. Blood samples were collected from patients visiting the North Colombo Teaching Hospital, Ragama, Sri Lanka who were clinically diagnosed by the consulting physician with either dengue fever (DF) or DHF. Blood was drawn within the first 5 days after fever onset.

Virus isolation, serotyping, and sequencing. The blood samples collected were confirmed as DENV-positive by reverse transcription polymerase chain reaction (RT-PCR) as previously described.[15] Virus isolation was done using the *Ae. albopictus* mosquito (C6/36) cell line. Cells (~2.5 million cells) were inoculated with 15 μL RT-PCR–confirmed DENV-positive serum for 1 hour. Cells were returned to fresh media, adjusted to appropriate pH using 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer, and incubated at 28°C without CO_2 for up to 12 days or until cells began to lift from plates (interpreted as cytopathic effect of infection). Supernatant containing viral particles was harvested and stored at −70°C. RNA was extracted from supernatant using the SV Total RNA isolation system (Promega, Madison, WI),

*Address correspondence to Aruna Dharshan De Silva, Genetech Research Institute, 54 Kitulwatte Road, Colombo 0800, Sri Lanka. E-mail: dslv90@yahoo.com

and DENV serotype was assigned using serotype-specific RT- PCR as previously described.[16]

Three viral isolates from 2012 were selected for full-genome sequencing. A fourth isolate from 2012 was partially sequenced along with two viruses from our repository isolated in 2003/2004.[9] The full genome was amplified for sequencing in 11 overlapping fragments using PCR primers listed in Supplemental Table 1 and illustrated in Supplemental Figure 1. Oligonucleotide primers for RT-PCR amplification and sequencing were designed using the PrimerSelect program from Lasergene (http://www.dnastar.com/t-primerselect.aspx). Reverse transcription was conducted separately for each fragment using M-MLV Reverse Transcriptase (Promega): 4 μL RNA was pre-annealed with 4 μM appropriate reverse primer at 65°C for 5 minutes, transferred to ice, and then reverse-transcribed in a total volume of 20 μL per product manual with 1 U/μL Rnasin ribonuclease inhibitor (Promega) for 45 minutes at 42°C. RT was heat-inactivated for 10 minutes at 70°C. Forward primer was added, and PCR was conducted using GoTaq Flexi DNA Polymerase (Promega) per the product instructions, with total $MgCl_2$ concentration of 3 mM and the following amplification: 5 minutes at 94°C followed by 35 cycles of 30 seconds at 94°C, 30 seconds at 53°C, and 1.5 minutes at 72°C and finally, 10 minutes at 72°C. Sequencing was by the Sanger method. Using the primers listed in Supplemental Table 1 for amplification and sequencing, we obtained at least single coverage and predominantly dual coverage over the entire genome. Sequencing was carried out at Macrogen (Seoul, South Korea) and the University of Pennsylvania DNA Sequencing facility. The amplification and sequencing approach is illustrated in Supplemental Figure 1.

Phylogenetic trees. Sanger sequences were assembled into contigs using the Sequencher sequence analysis software (version 5.1; Gene Codes Corporation). A consensus sequence was generated for the full genome and the envelope (E) gene from the three fully sequenced 2012 Sri Lankan viruses using the Macvector software (version 12.0.3; Accelrys). These consensus sequences were used to search the nucleotide (nr/nt) collection in Genbank for closely related sequences using the National Center for Biotechnology Information BLAST tool (megablast).[17,18] The 100 most similar sequences for each consensus were retrieved for phylogenetic analyses. We omitted sequences for which we could not determine country of origin. Alignments were performed using the ClustalW algorithm within the Macvector software (version 12.0.3; Accelrys) or for trees, the Needleman–Wunsch algorithm (implemented within the R programming package) with matches scoring one and gaps and mismatches scoring zero to minimize Levenshtein distance between sequences and consensus.[19]

Phylogenetic relationships between strains were investigated using the BEAST software package (version 1.7.5),[20] which implements the Bayesian Markov Chain Monte Carlo (MCMC) method.[21] All trees were constructed taking into consideration time of isolation and using a GTR + Γ_4 + I model of nucleotide substitution with three codon positions and substitution, rate heterogeneity, and base frequencies unlinked across all codon positions (as has been shown previously to be appropriate).[22,23] Four independent MCMC samples of 10 million iterations each were performed for each set of sequences and model. Samples were recorded every 1,000 generations and combined with 10% burn-in removal using the Log Combiner program (implemented in BEAST). Convergence of the chain and sufficient effective sampling size (ESS; ESS > 200) were confirmed using the Tracer program.[24] The geographic locations of ancestral (node) states and migration patterns were examined using Bayesian Stochastic Search Variable Selection (BSSVS) implemented in BEAST. Trees constructed assuming relaxed (uncorrelated log normal) and strict molecular clock models showed similar topology and coalescent times (relaxed clock analysis not shown; available on request).

For trees using full-genome sequences, we chose a model assuming a uniform prior distribution of rates of all possible substitutions. The assumption of log-normal distributions of these features better fit the E gene data. Geographic analysis was performed assuming both symmetric and asymmetric migration rates between locations as well as uniform and log-normal prior distributions of migration rates with largely similar results (asymmetric results available on request). Maximum clade credibility trees were generated using TreeAnnotator (part of BEAST) and visualized in FigTree.[25] We adapted trees for visualization on Google Earth (http://www.google.com/earth/index.html) using the SPREAD program (http://www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html).

Recombination and selection pressure analysis. We performed recombination and selection analyses on the open reading frame (ORF) of DENV strains using the hypothesis testing using phylogenes (HyPhy) package hosted on the Datamonkey webserver (www.datmonkey.org) or locally (for single breakpoint recombination over the ORF).[26,27] In addition, we partitioned the viral genomic sequence dataset into each of 10 protein genes of the DENV genome for analysis. Models were selected using HKY as the nucleotide substitution model for analyses on the ORF dataset, whereas the TRN-93 model was used for individual gene datasets.

Single breakpoint recombination (SBR) was not detected in either the full genome dataset or in each of the ORFs by the HyPhy software package (no improvement in Akaike information criterion [AIC] or Bayesian information criterion [BIC]). The dN/dS ratios (omega) were calculated using codon-based maximum likelihood approaches: single likelihood ancestor (SLAC), fixed effects likelihood (FEL), and the internal branch fixed effects likelihood (IFEL). The random effects likelihood method (REL) was used to calculate dN/dS for smaller protein genes because of alignment size restrictions on the DataMonkey server (not used for ORF, viral E, or NS5 genes). Alignment-wide dN/dS ratios were calculated using the SLAC approach.[28,29] The mixed effects model of evolution method (MEME) was also used to identify both diversifying and episodic (affecting only a subset of lineages) selection in the datasets.[30] Codons were reported as under positive or negative selection if P value was less than or equal to 0.1 for SLAC, FEL, MEME, and IFEL methods. Codons were reported as under selection using the REL method with a Bayes factor cutoff of 50.

Mutational analysis. Mutation rate was estimated using the BEAST package with the models indicated above under strict molecular clock assumption. To identify mutations that collected in the circulating Sri Lankan DENV-1 strain between 2009 and 2012, we compared each of the sequenced 2012 viruses with the consensus sequence of five available DENV-1 genomes collected in Sri Lanka in 2009/2010 and each of these viruses individually (Genbank accession numbers JN054256, HQ891315, HQ891314, HQ891313, and

JN054255). This comparison allowed us to find novel mutations that had occurred in the genome as well as variations that had been present at low levels in 2009 and persisted in 2012 viruses. For analysis of distribution of mutations, ambiguous bases in the sequences were omitted to eliminate the confounding effect of decreased sequence quality at ends of reads. For analysis of non-synonymous changes, we determined sequence identity for selected related DENV-1 isolates at sites of variation between 2009 and 2012 Sri Lankan viruses. Analyses were performed in R version 3.0.1.[31]

Data access. Genome sequences have been uploaded to GenBank with the following names and accession numbers: SL_2012_GS0289, GenBank ID KJ726663; SL_2012_GS0292, GenBank ID KJ726665; SL_2012_GS0308, GenBank ID KJ726664; SL_2012_GS0319, GenBank ID KJ726662.

RESULTS

Serotype of predominant disease-causing DENV in 2012. The incidence of dengue disease was high in Sri Lanka in 2012, peaking in the capital city of Colombo in July, which it had in the previous 3 years. Over 1,000 cases were reported per month in the city from June to August of 2012.[13] To assess which serotype of DENV was responsible for severe dengue disease during the Colombo epidemic of 2012, we collected serum of patients clinically diagnosed with severe DF or DHF and hospitalized in the Sri Lankan Government Teaching Hospital, Colombo North (Ragama) during the first 4 months of 2012. Of 19 tested serum samples, 15 samples were determined to contain DENV-1, and these viruses were expanded for additional study. In the remaining four samples, serotyping was not completed because of low viral load. No other serotypes were found, although other recent studies have observed DENV-4 from 2012 onwards (unpublished data).

Full-genome sequencing of 2012 Sri Lankan DENV-1 and phylogeography. To determine the origins of the DENV-1 responsible for severe dengue disease in Colombo in 2012, we sequenced the full genomes of three of the collected 2012 viruses (GS0289, GS0308, and GS0319). Using Bayesian methods described above, we conducted phylogenetic analyses, comparing these genomes with a collection of previously published full-length DENV-1 sequences: the 100 DENV-1 genomes on GenBank most similar (smallest edit distance) to the 2012 viruses as well as representative DENV-1 isolates of different genotypes, including virus isolated in nearby Kerala, India in 2009 (Figure 1).[32] All three 2012 viruses belonged to the DENV-1 genotype I, which is commonly found in southeast Asia (Figure 1A). The most similar recovered sequences were those sequences of several DENV-1 clones recovered in Sri Lanka during 2009/2010, lending support to the idea that the 2012 DENV-1 viruses sequenced were descended from viruses circulating in Sri Lanka in 2009 (Figure 1B). The 2012 viruses shared 99.41–99.68% identity with the available Sri Lankan viruses from 2009/2010. Moreover, our phylogeographic analysis suggests that the 2012 viruses evolved from a Sri Lankan ancestral virus (probability approaches 1.0) and that the common Sri Lankan ancestral virus from which the 2009 and 2012 viruses stemmed was present on the island 4.59 years before 2012 (95% highest posterior density interval [HPD] = 4.18–5.01 year) or in 2007.

Phylogeographic analysis additionally provided clues as to the route that this virus took to Sri Lanka. The Sri Lankan viruses seem to have derived from a virus that caused a severe epidemic in southern China in 2006 (posterior probability = 0.80) (Figure 1B).[33] This virus, in turn, likely derived from a virus in Thailand (posterior probability = 0.91), which also gave rise to a strain that was responsible for severe dengue disease in Thailand in 2001.[34] Our data suggest that a closely related virus spread in Thailand, Cambodia, and Vietnam since 2001. We note that the 2012 Sri Lankan virus is not closely related to the etiologic strain of an outbreak of DHF in nearby Kerala, India in 2009. This Indian virus belongs to genotype III, which has been previously shown (Figure 1A).[32] Supplemental Video 1 illustrates the routes of dissemination of the DENV-1 genotype I strain as inferred by Bayesian phylogeographic analysis.

E sequence of 2012 Sri Lankan DENV-1. Although full-genome sequencing characterizes all sites of variation in the genome, enabling precise analysis, phylogenetic analyses of DENV have traditionally relied on segments of the viral genome, such as the E gene, and therefore, many more partial genome sequences are available in Genbank. If the virus traveled to Sri Lanka by way of an additional intermediate country where full-length sequencing was not performed, we would miss this step in the above analysis. To improve our resolution, we narrowed our database queries to the E, again recovering the 100 sequences most similar to the 2012 Sri Lankan consensus. We used these sequences and other selected DENV-1 sequences to generate a second phylogeographic tree (Figure 2 and Supplemental Video 2).

The E-based analysis confirmed that the 2012 viral strain belonged to the same clade as those strains that caused the 2009 epidemic. Consistent with the full-genome analysis, this phylogenetic tree also shows that the Sri Lankan viruses derived from a Chinese ancestral virus. The tree illustrates extensive interchange between Thailand and China, and it implies that the China/Sri Lanka clade evolved from viruses circulating in Thailand in 2001 (although posterior probabilities supporting it are low).

Tracing the Sri Lankan strain's path forward in time, it seems that it may have spread from Sri Lanka to Pakistan, which was noted previously,[35] and Singapore. The phylogeographic analysis suggests that the transmission to Singapore may have occurred two times in independent events; however, the statistical support for this topology is weak, which is evidenced by low posterior values at Sri Lankan and Singaporean ancestral nodes in Figure 2. We also sequenced the E gene of two DENV-1 isolates present in Sri Lanka before 2009 (in 2003 or 2004).[9] Phylogenetic analysis confirmed that these viruses belonged to genotype IV and did not likely give rise to the 2009–2012 epidemic strain (Supplemental Figure 2).[9]

Recombination and selection pressures. It is possible that the presence of recombinant DENV-1 genomes in our datasets might yield erroneous phylogenetic topologies. To exclude this possibility, we screened out datasets using SBR analysis across all viruses. We found no evidence of recombination within the viral ORF or within each individual protein-coding gene. The tree in Figure 1, constructed from full-length genomes, showed similar topology and divergence times as trees constructed using the isolated coding regions of the matrix protein and non-structural protein-3 of the same viruses (data not shown). This finding again suggests that recombination in these regions did not bias the phylogenetic analyses.

We also investigated our dataset of viral genomes for evidence of selection pressures. The dN/dS (ω) ratio across the

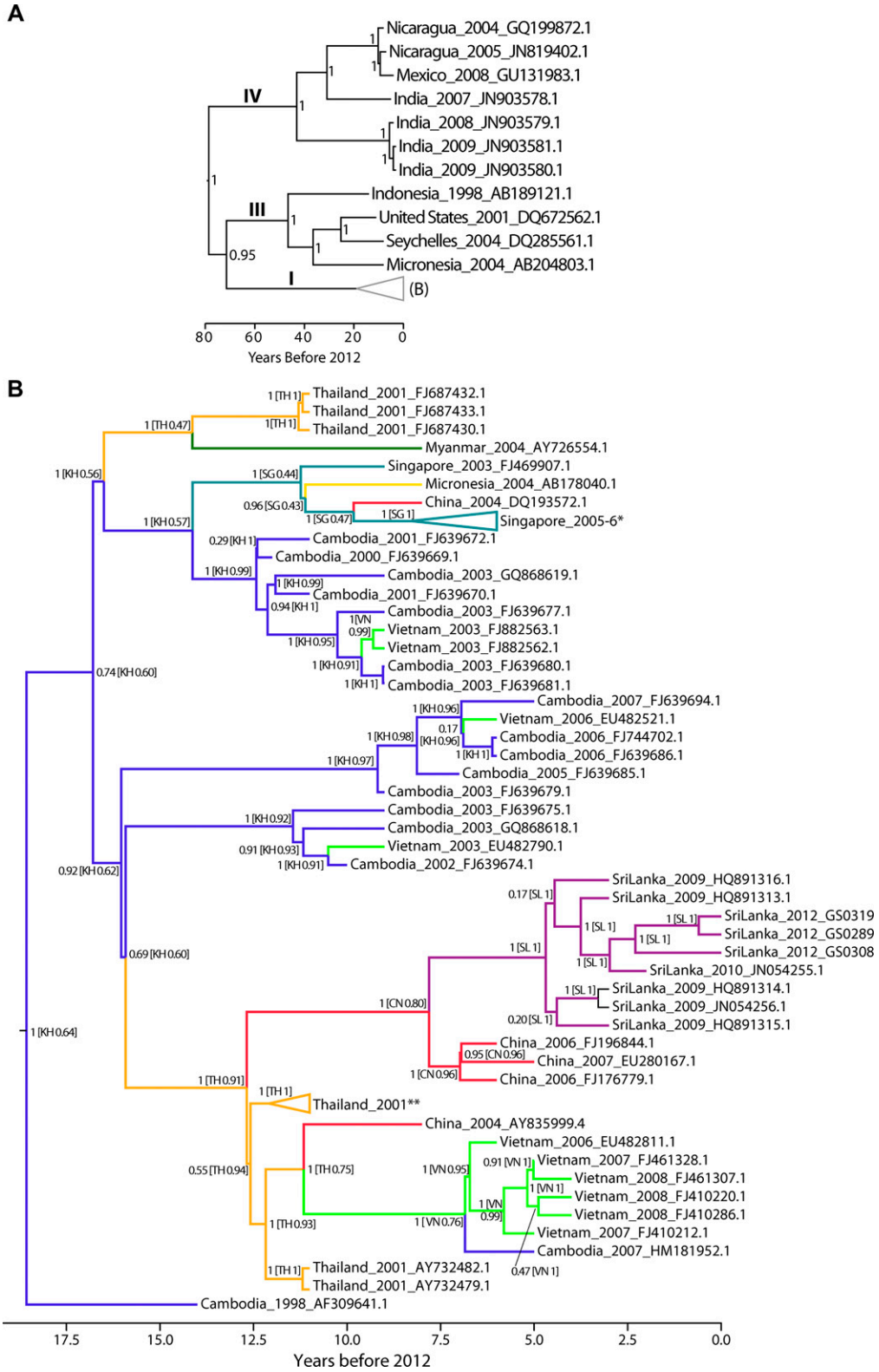

FIGURE 1. Phylogenetic relationships of complete genomes of DENV-1. (**A**) Maximum clade credibility tree using representative sequences from major genotypes of DENV-1, 3 full genomes from Colombo, Sri Lanka in 2012, and the 100 most similar sequences available in Genbank with strict clock. Genotype is indicated on branches. Viruses in genotype I are shown in **B**. The isolate names indicate country of origin, year of sampling, and GenBank accession number. (**B**) Maximum clade credibility tree showing phylogeography of full genomes of DENV-1 genotype I isolates from **A**, assuming symmetric migration rates between locations. Posterior support is indicated at each node. Ancestral states and their probabilities are indicated in brackets. Sequences at collapsed branches (GenBank accession numbers): *EU081226.1–EU081234.1, EU081236.1–EU081256.1, EU081259.1, EU081261.1, EU081263.1–EU081275.1, and EU081278.1–EU081280.1; **FJ687428.1, FJ687431.1, FJ687429.1, FJ687426.1, FJ687427.1, and FJ850068.1. CN = China; KH = Cambodia; SG = Singapore; SL = Sri Lanka; TH = Thailand; VN = Vietnam.

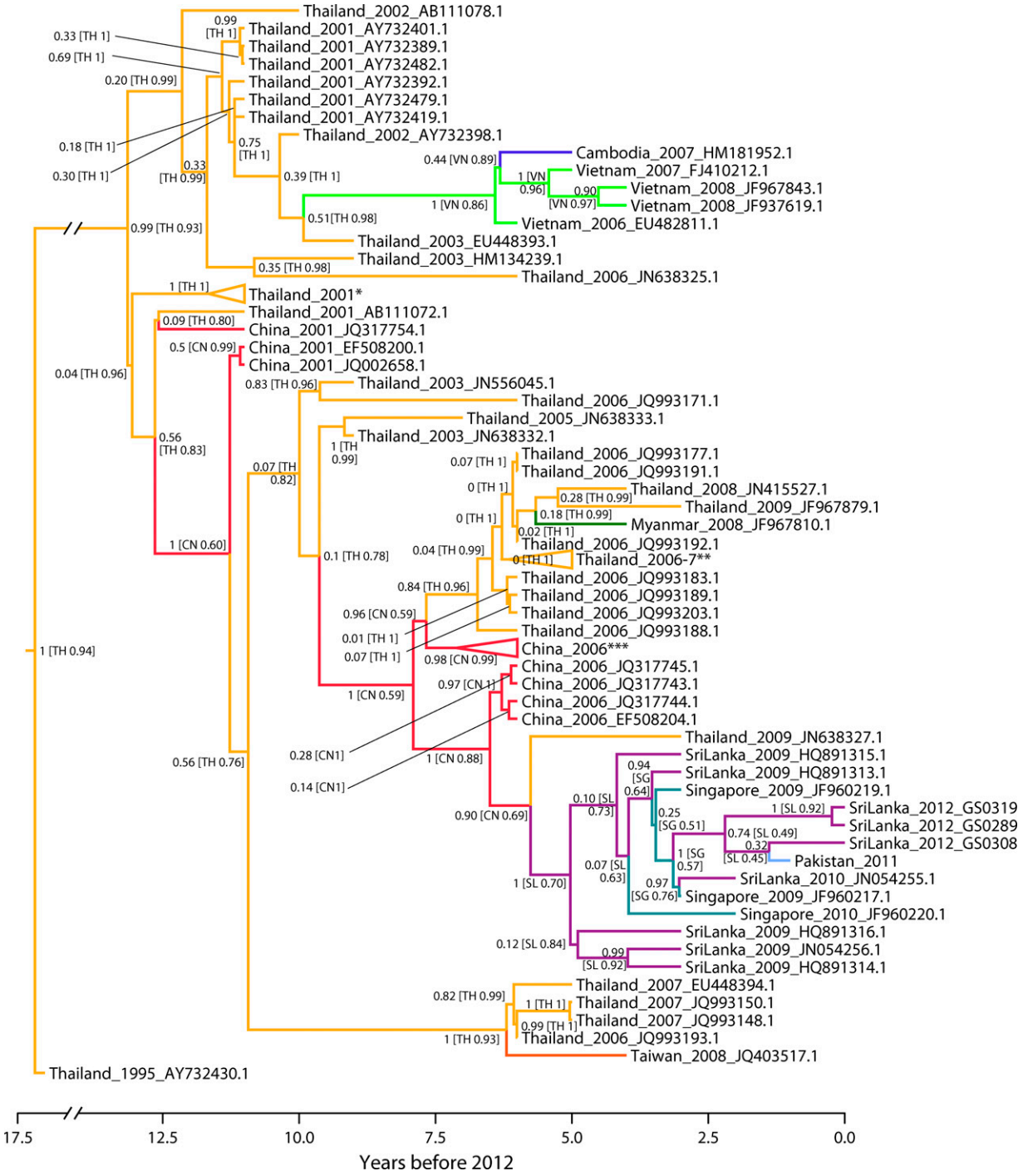

FIGURE 2. Phylogenetic relationships of E sequences of DENV-1. Maximum clade credibility tree incorporating geographic data for the E gene of the 2012 Sri Lankan DENV-1 isolates, the 100 most similar sequences obtained from Genbank, and the E sequence of a virus isolated in Pakistan in 2011. The model incorporates strict clock and symmetric migration rates. Posterior support is indicated at each node. Ancestral states and their probabilities are indicated in brackets. Sequences at collapsed branches (GenBank accession numbers): *FJ687426.1–FJ687429.1, FJ687431.1, EU117304.1–EU117308.1, EU117312.1, and FJ850068.1; **JQ993108.1, JQ993127.1, JQ993130.1, JQ993132.1, JQ993133.1, JQ993184.1, and JQ993198.1; ***EF508205.1, EF113152.1, EF113153.1, FJ176779.1, FJ196844.1, FJ196855.1–FJ196860.1, JQ277849.1, JQ277850.1, JQ277853.1, JQ277854.1, and JQ277864.1–JQ277874.1.

viral ORF in our dataset of all viruses used in the above phylogenetic trees was 0.0640 using the single likelihood ancestor (SLAC) method, suggesting that DENV-1 is under strong purifying selection, which has been previously observed for other DENV strains and arboviruses.[36–38] Gene-by-gene analysis also revealed strong purifying selection within each protein-coding gene (Table 1). We discuss the effect that it might have had on our phylogenetic trees below.

By contrast, codons under positive selection were relatively infrequent, with only eight codons identified by at least two (of four) methods of detection in analysis of the viral ORF (capsid, C-66; matrix, M-93; non-structural protein 1 [NS1],

TABLE 1
Selection pressure analysis

Gene*	dN/dS (SLAC method)	No. of codons under positive selection†					No. of codons under negative selection‡			
		≥ 1 method	≥ 2 methods	≥ 3 methods	≥ 4 methods	≥ 5 methods	≥ 1 method	≥ 2 methods	≥ 3 methods	≥ 4 methods
C	0.173	0	0	0	0	0	41	21	11	6
M	0.0871	0	0	0	0	0	92	52	35	15
E	0.0681	7	0	0	0	NA	161	117	59	NA
NS1	0.0907	4	1	1	0	0	123	102	68	40
NS2A	0.115	7	1	1	0	0	93	51	32	6
NS2B	0.0483	1	0	0	0	0	45	38	14	0
NS3	0.0411	7	2	0	0	0	215	173	141	69
NS4A	0.0650	5	1	1	1	0	87	46	23	6
NS4B	0.0357	1	0	0	0	0	85	66	57	24
NS5	0.0685	14	1	1	0	NA	285	207	95	NA
ORF	0.0640	56	8	3	0	NA	1,044	740	496	NA

Cutoff *P* value for reporting of codons under positive or negative selection was 0.1 for SLAC, FEL, MEME, and IFEL methods. Cutoff Bayes factor for codons detected using REL method was 50. NA indicates that the dataset was analyzed with only four methods.

*Dataset of 114 full-length viral genome sequences partitioned into individual protein-coding genes or the viral ORF for analysis.

†Reported are the numbers of codons predicted to be under positive selection by at least one method tested. All datasets were analyzed for positive selection using codon-based maximum likelihood approaches (SLAC, FEL, and IFEL) and the branch site method (MEME). Smaller datasets were analyzed additionally with REL.

‡All datasets were analyzed for negative selection with SLAC, FEL, and IFEL methods, and smaller datasets were analyzed with REL method.

NS1-94; NS2A-159; NS3-164; NS3-589; NS4A-94; NS5-379) (Table 1). In gene-by-gene analysis, only the protein-coding regions of NS1, NS2A, NS4A, and NS5 contained codons identified to be under positive selection by at least three methods (of five methods used; four methods used for E and NS5 genes; data available on request) (Materials and Methods and Table 1). Analysis by the most inclusive of the models, MEME, a branch site method that detects diversifying selection affecting only a subset of lineages, identified a larger number of sites under positive selection, suggesting that the DENV-1 genome may be subject to evolution by episodic selection.[30] However, even this method revealed no positively selected codons in the E and M genes (data available on request).

Mutational and evolutionary analysis. Our phylogenetic analysis suggests that the DENV strain that caused the 2009 outbreak persisted in Sri Lanka and continued to cause disease through 2012. Dengue disease levels were also high in Sri Lanka during 2010 and 2011,[13] and whereas we have limited genomic data for these years, published and unpublished data suggest that DENV-1 was responsible for much of the disease in these years (Sirisena[14] and A. D. De Silva, unpublished data). Thus, it is conceivable that the 2012 strain caused and evolved over four consecutive seasonal outbreaks of dengue disease in Sri Lanka. Based on our phylogenetic analysis of full-length viruses, we estimated an overall mutation rate of 9.0×10^{-4} substitutions/site per year (95% HPD = $8.0 \times 10^{-4} - 1.0 \times 10^{-3}$ substitutions/site per year), roughly agreeing with previous estimations.[39,40] Considering just the Sri Lankan outbreak viruses, we estimate a similar mutation rate of 9.4×10^{-4} substitutions/site per year (95% HPD = $6.4 \times 10^{-4} - 1.3 \times 10^{-3}$ substitutions/site per year).

To further investigate the evolutionary process of the 2012 viruses from the 2009 ancestral virus, we compared our three complete 2012 DENV-1 genomes as well as a partial sequence of a fourth 2012 isolate (GS0292) from Sri Lanka with the five available sequences of Sri Lankan 2009/2010 DENV-1 isolates. Novel synonymous and non-synonymous mutations accumulated within the 2012 viruses across the entire DENV genome, with no obvious propensity for specific regions (Figure 3A). As expected, most of these novel mutations were synonymous.

To determine whether any amino acid changes may have become fixed in the viral population since 2009, we compared the 2012 DENV-1 sequences with a consensus sequence of the 2009 isolates. We identified 16 nucleotides at which a 2012 virus contained a non-synonymous difference (Figure 3B). Some of these mutations were already present (but not consensus) in a minority of the 2009 viruses. Non-synonymous mutations in four codons were shared by at least two of the 2012 isolates, suggesting possible fixation within the 2012 DENV-1 population (although sequencing of additional 2012 isolates is required to determine it). They occurred at nucleotides 1749 (E-T272M), 3043 (NS1-E208D), 4563 (NS3-R15K), and 6917 (NS4B-V31I). Of these four nucleotides, only the NS1 mutation is novel. The E and NS3 variants were each present in one Sri Lankan 2009 virus and the 2010 Sri Lankan isolate, and NS4B-V31I was present in the 2010 isolate. None of these four sites were identified as being under selection in above selection analyses.

The four potentially fixed amino acid changes are predominantly conservative and unlikely to significantly alter the structure or function of their respective proteins. The NS3-R15K mutation falls within the protease domain of the NS3, but it is located in a poorly conserved portion,[41] and it is predicted to be distant from the catalytic site and specificity residues.[42] The conservative NS1-E208D change is probably also insignificant, although little detailed structural information is available for NS1.[43] The NS4B protein is an endoplasmic reticulum (ER)-localized transmembrane protein that is not well-studied in DENV-1; however, in DENV-2, the luminal residue equivalent to V31 is an isoleucine, and therefore, the conservative V31I change in NS4B is unlikely to alter function.[44] Finally, based on the published structure of the DENV-2 E,[45] E-T272M falls within a small helix that rearranges on membrane fusion to facilitate presentation of the hydrophobic fusion loops. The change from polar to hydrophobic residue may potentially alter the dynamics of this step. However, the equivalent residue in DENV serotype 2 is a methionine; therefore, the T272M mutation may not alter function.[45]

DISCUSSION

The 2009 dengue disease epidemic in Sri Lanka is thought to be caused by the arrival of a new DENV-1 of genotype I. Our data suggest that the DENV-1 causing the bulk of serious dengue disease in and near the capital city of

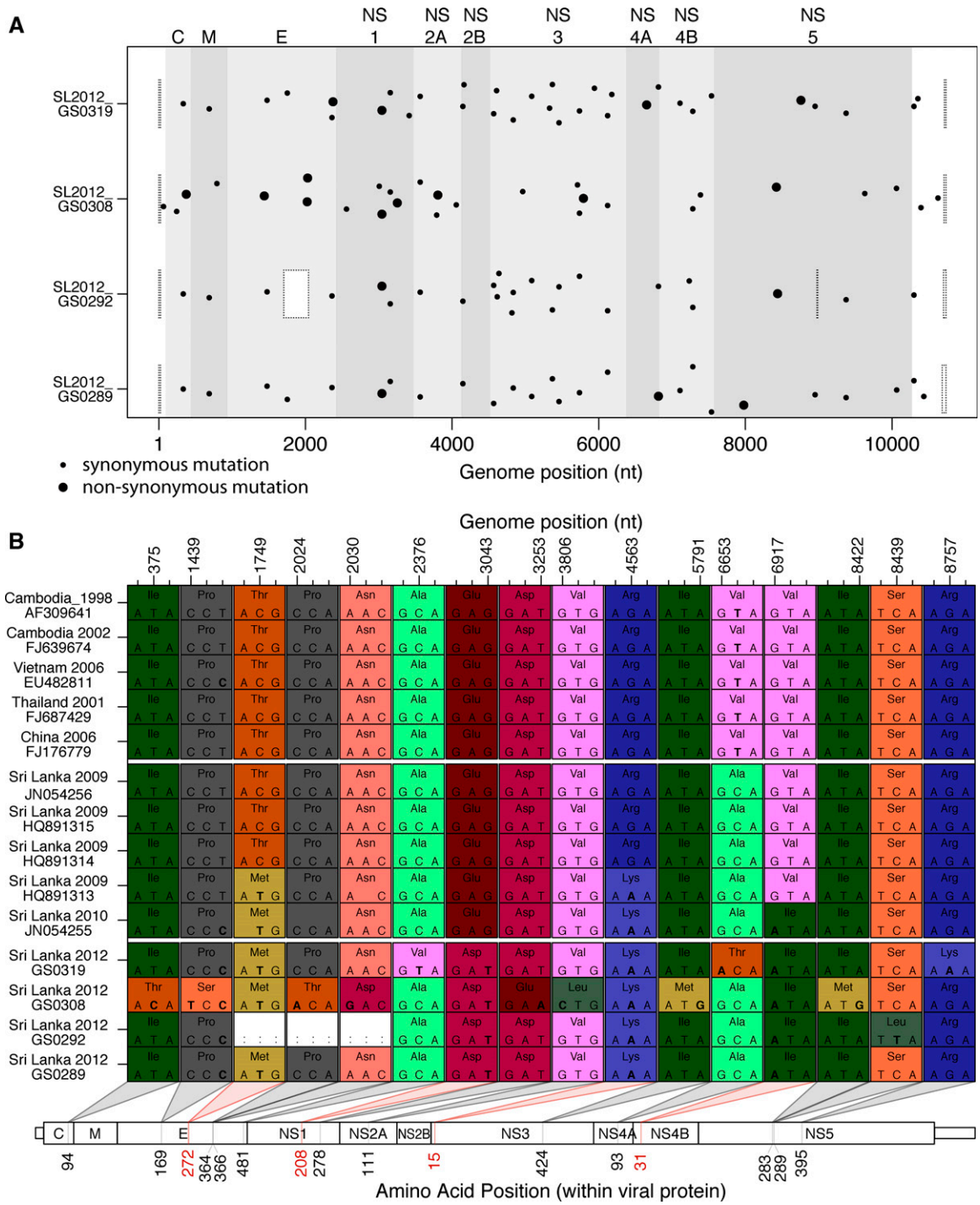

FIGURE 3. Evolution of Sri Lankan DENV-1 between 2009 and 2012. (**A**) Circles indicate novel mutations in each of four sequenced 2012 DENV-1 isolates not present in any of five available 2009/2010 sequences (Genbank accession numbers JN054256, HQ891315, HQ891314, HQ891313, and JN054255). Mutations are plotted at their nucleotide locations as indicated at the bottom. Gene boundaries are indicated by alternating shaded regions, and gene identities are indicated at the top. Small circles indicate synonymous mutations; large circles indicate non-synonymous mutations. Dashed rectangles indicate regions of the genome that were not sequenced. (**B**) The aligned polyproteins of the 2012 Sri Lankan viruses were compared with the consensus sequence of the 2009/2010 viruses. Positions not completely conserved are shown with amino acid and codon. At each position, data are provided for the 2009/2010 Sri Lankan viruses and selected isolates from the strain's phylogenetic heritage for comparison. Genome nucleotide coordinate and protein location are indicated at the top and bottom, respectively. Changes occurring in at least two 2012 viruses are labeled in red at bottom.

Colombo in 2012 was descended from a Sri Lankan virus that also gave rise to the 2009 genotype I virus. We infer from this information that the same DENV-1 strain has likely caused 4 years of repeated epidemics in Colombo and surrounding areas. The extended success of a single strain is surprising, and it suggests that there was little immunity to this strain of DENV-1 in the Sri Lankan population. Notably, DENV-1 of genotype IV, which circulates

mainly in the Americas, has circulated in Sri Lanka as recently as 2004.[9] However, it did not cause similar widespread severe dengue disease.

Previous studies have observed similar phenomena. In the 1980s, the arrival of a southeast Asian genotype of DENV-2 in Cuba and then South and Central America caused DHF epidemics, despite established cocirculation of DENV-2 as well as DENV-1 and DENV-3 in the affected countries. Before this time, DHF had been rare in the region, suggesting the existence of viral determinants of disease severity.[46,47] Virulence motifs were later identified in the E gene and within the 3′ and 5′ untranslated regions (UTRs) of DENV-2.[48,49] However, a DENV-1 genotype IV strain that caused extensive DHF in French Polynesia spread to Hawaii, where it caused only DF without severe disease. Comparison of viral genomes isolated in French Polynesia and Hawaii revealed no potential changes to virulence factors, implying that there are important roles for host factors—human immune determinants (inherited or acquired during previous exposures) and mosquito vector differences.[50]

The genotype IV strain of DENV-1 circulating in Sri Lanka in 2003/2004 did not spread broadly within the population, which could be attributed to a lack of virulence or fitness factors in the virus, differences in mosquito populations, or less likely, changes in the herd immunity among Sri Lankans between 2003 and 2009. In any case, it is unlikely that DENV-1 cross-protective immunity existed within the population in 2009. Whether the increased virulence of the 2009–2012 genotype I virus is intrinsic to the strain or can be attributed to immunologic determinants found within the Sri Lankan population is unclear. Given that related and ancestral viruses caused similar epidemics across south Asia, we speculate that the former case is at least partially true. Additional work is underway to compare the full genomes of the DEN-1 viruses isolated in Sri Lanka in 2003/2004 (genotype IV) with the 2009–2012 genotype I viruses to address this question.

Phylogeographic analysis of the Sri Lankan DENV-1 isolates from 2009 to 2012 using both full-genome sequences and E gene sequences suggests that the genotype I virus was spread, directly or indirectly, from Thailand to China to Sri Lanka. The virus may have arrived in Sri Lanka through intermediate locations where sequencing was not performed, which are, therefore, undetectable to us. However, it is evident that there has been extensive migration of DENV-1 throughout south and southeast Asia, especially between Thailand and China in the past decade. Estimation of divergence times (or time to most recent common ancestor [tMRCA]) suggests that currently circulating Sri Lankan DENV-1 descends from a virus present in Sri Lanka by 2007. In that year, two major construction projects led by Chinese contractors began in Sri Lanka, bringing hundreds of Chinese workers through Colombo.[51,52] Whether these events relate to the spread of the genotype I virus to Sri Lanka cannot be known; however, they do illustrate the increasing globalization of southeastern Asia, which has, in part, enabled the resurgence of severe dengue disease.[7,8] Direct flights now link most countries in the region, making containment of such virulent arboviruses difficult. Notably, the descendants of the ancestral virus have been responsible for several epidemics across southeastern Asia,[33,39,53] suggesting little immunity to this genotype in the region and/or relative virulence of the strain. To this end, our analysis also confirms the recent spread of the virus to Pakistan[35] and exchange of the virus between Sri Lanka and Singapore. At least two daily flights connected Singapore and Sri Lanka throughout most of the last decade, with four daily flights in recent times, setting the stage for ongoing spread of DENV between the two countries.

We acknowledge that strong purifying selection pressures, evidence of which we detected in DENV-1, have been shown to potentially skew estimates of divergence time in phylogenetic analyses.[54] Although there may be an effect of purifying selection on our calculated divergence times, we note that the timing calculated for our phylogeographic analyses correspond well with the above noted events affecting immigration patterns to Sri Lanka. This correlation lends credence to the reported divergence times. In addition, trees made with different sections of the genome (under slightly different selection pressures) support similar conclusions.

Our analysis of DENV-1 across our phylogenies and within Sri Lanka from 2009 to 2012 shows an evolutionary rate on the order of 10^{-3} substitutions/site per year, consistent with previous reports.[39,40,55] We saw no evidence for a considerable difference in the evolutionary rate of the virus during the outbreaks in Sri Lanka. Four presumably conservative non-synonymous mutations that were not consensus in 2009 accumulated in the 2012 viruses. Additional sequencing work will be required to determine whether any of the observed mutations were under positive selection. Regardless of their impact on protein structures and functions, each of these mutations has the potential to alter immune epitopes, and additional work is underway to study their effects on viral visibility to T cells in the human host.

Work is ongoing to determine why this particular strain has caused so much severe dengue disease in Sri Lanka—the answer is likely to involve a combination of host immune factors and virus-specific factors. The widespread severity of disease caused by this viral strain also illustrates the need for better international control efforts and the development of treatments or vaccines. Studies such as our study will help to direct the allocation of such materials should they become available.

Received September 10, 2013. Accepted for publication March 24, 2014.

Published online May 5, 2014.

Note: Supplemental table, figures, and videos appear at www.ajtmh.org.

Acknowledgments: The authors thank Frederic Bushman and Robert Doms at the University of Pennsylvania for their advice and generous help with reagents and sequencing efforts. We thank Dr. Aravinda de Silva for provision of the 2003/2004 DENV-1 strains and materials for sequencing those strains as well as valuable intellectual conversation. The staff and research assistants at Ward 9, North Colombo Teaching Hospital, Ragama are acknowledged for the help in providing samples, and the authors thank the patients for willingly providing samples. Finally, the authors thank the staff at Genetech for constructive discussion and technical guidance.

Financial support: Funding for this work was provided by National Institutes of Health Contract Grant HHSN272200900042C through a subcontract awarded to A.D.D.S; K.E.O. was supported by the American Society of Tropical Medicine and Hygiene Benjamin H. Kean Traveling Fellowship in Tropical Medicine and the James S. Porterfield Prize in International Virology.

Authors' addresses: Karen E. Ocwieja and Scott Sherrill-Mix, Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, E-mails: kocwieja@mail

.med.upenn.edu and shescott@mail.med.upenn.edu. Anira N. Fernando, Rashika N. Tennekoon, Rashmi Tippalagama, Shivankari Krishnananthasivam, and Aruna Dharshan De Silva, Genetech Research Institute, Colombo, Sri Lanka, E-mails: anira@genetechsrilanka .com, rashika@genetechsrilanka.com, rashmi.tippalagama@gmail.com, shivanky@yahoo.com, and dslv90@yahoo.com. Sesh A. Sundararaman, Division of Hematology/Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, E-mail: sesh@mail.med .upenn.edu. Gayani Premawansa, North Colombo Teaching Hospital, North Colombo, Sri Lanka, E-mail: gavisprema@gmail.com. Sunil Premawansa, Department of Zoology, University of Colombo, Colombo, Sri Lanka, E-mail: suviprema@gmail.com.

REFERENCES

1. Nathan MB, D-D R, Guzman M, 2009. *Dengue: Guidelines for Diagnosis, Treatment, Prevention, and Control: New Edition*. Geneva: World Health Organization.
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, Myers MF, George DB, Jaenisch T, Wint GR, Simmons CP, Scott TW, Farrar JJ, Hay SI, 2013. The global distribution and burden of dengue. *Nature 496:* 504–507.
3. Guzman MG, Kouri G, 2002. Dengue: an update. *Lancet Infect Dis 2:* 33–42.
4. Halstead SB, et al., 1970. Observations related to pathogenesis of dengue hemorrhagic fever. I. Experience with classification of dengue viruses. *Yale J Biol Med 42:* 261–275.
5. Rothman AL, 2011. Immunity to dengue virus: a tale of original antigenic sin and tropical cytokine storms. *Nat Rev Immunol 11:* 532–543.
6. Whitehorn J, Simmons CP, 2011. The pathogenesis of dengue. *Vaccine 29:* 7221–7228.
7. Gubler DJ, 2002. The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res 33:* 330–342.
8. Wilder-Smith A, Gubler DJ, 2008. Geographic expansion of dengue: the impact of international travel. *Med Clin North Am 92:* 1377–1390.
9. Kanakaratne N, et al., 2009. Severe dengue epidemics in Sri Lanka, 2003–2006. *Emerg Infect Dis 15:* 192–199.
10. Messer WB, Gubler DJ, Harris E, Sivananthan K, de Silva AM, 2003. Emergence and global spread of a dengue serotype 3, subtype III virus. *Emerg Infect Dis 9:* 800–809.
11. Tissera HA, et al., 2011. New dengue virus type 1 genotype in Colombo, Sri Lanka. *Emerg Infect Dis 17:* 2053–2055.
12. Dissanayake VH, Gunawardena ND, Gunasekara NC, Siriwardhana DR, Senarath N, 2011. Shift in the transmission pattern of dengue serotypes and concurrent infection with more than one dengue virus serotype. *Ceylon Med J 56:* 176–178.
13. Sri Lankan Ministry of Health, EU, 2013. *Disease Surveillance, Trends*. Available at: http://www.epid.gov.lk/web/index.php?option=com_casesanddeaths&Itemid=448&lang=en. Accessed June 10, 2013.
14. Sirisena PD, Noordeen F, 2014. Evolution of dengue in Sri Lanka-changes in the virus, vector, and climate. *International journal of infectious diseases: IJID: Official Publication of the International Society for Infectious Diseases 19:* 6–12.
15. Sudiro TM, et al., 1997. Rapid diagnosis of dengue viremia by reverse transcriptase-polymerase chain reaction using 3′-noncoding region universal primers. *Am J Trop Med Hyg 56:* 424–429.
16. Lanciotti RS, Calisher CH, Gubler DJ, Chang GJ, Vorndam AV, 1992. Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *J Clin Microbiol 30:* 545–551.
17. Morgulis A, et al., 2008. Database indexing for production MegaBLAST searches. *Bioinformatics 24:* 1757–1764.
18. Zhang Z, Schwartz S, Wagner L, Miller W, 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol 7:* 203–214.
19. Needleman SB, Wunsch CD, 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol 48:* 443–453.
20. Drummond AJ, Suchard MA, Xie D, Rambaut A, 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol 29:* 1969–1973.
21. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W, 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics 161:* 1307–1320.
22. Dunham EJ, Holmes EC, 2007. Inferring the timescale of dengue virus evolution under realistic models of DNA substitution. *J Mol Evol 64:* 656–661.
23. Rabaa MA, Ty Hang VT, Wills B, Farrar J, Simmons CP, Holmes EC, 2010. Phylogeography of recently emerged DENV-2 in southern Viet Nam. *PLoS Negl Trop Dis 4:* e766.
24. Rambaut A, Drummond AJ, 2007. Tracer v1.5. Available at: http://tree.bio.ed.ac.uk/software/tracer/.
25. Morariu VI, Srinivasan BV, Raykar VC, Duraiswami R, Davis LS, 2008. Automatic online tuning for fast Gaussian summation. *Adv Neural Inf Process, 21 edn.* 1113–1120.
26. Pond SL, Frost SD, Muse SV, 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics 21:* 676–679.
27. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL, 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics 26:* 2455–2457.
28. Kosakovsky Pond SL, Frost SD, 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol 22:* 1208–1222.
29. Pond SL, et al., 2006. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLOS Comput Biol 2:* e62.
30. Murrell B, et al., 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet 8:* e1002764.
31. Core Team R, 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
32. Anoop M, et al., 2012. Complete genome sequencing and evolutionary analysis of dengue virus serotype 1 isolates from an outbreak in Kerala, South India. *Virus Genes 45:* 1–13.
33. Chen S, 2011. The origin of dengue viruses caused the DF outbreak in Guangdong province, China, in 2006. *Infect Genet Evol 11:* 1183–1187.
34. Zhang C, et al., 2005. Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. *J Virol 79:* 15123–15130.
35. Khan MA, et al., 2013. Emergence and diversification of dengue 2 cosmopolitan genotype in Pakistan, 2011. *PLoS ONE 8:* e56391.
36. Anez G, Morales-Betoulle ME, Rios M, 2011. Circulation of different lineages of dengue virus type 2 in Central America, their evolutionary time-scale and selection pressure analysis. *PLoS ONE 6:* e27459.
37. Holmes EC, 2003. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol 77:* 11296–11298.
38. Anez G, et al., 2013. Evolutionary dynamics of West Nile virus in the United States, 1999–2011: phylogeny, selection pressure and evolutionary time-scale analysis. *PLoS Negl Trop Dis 7:* e2245.
39. Schreiber MJ, et al., 2009. Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol 83:* 4163–4173.
40. Twiddy SS, Holmes EC, Rambaut A, 2003. Inferring the rate and time-scale of dengue virus evolution. *Mol Biol Evol 20:* 122–129.
41. Valle RP, Falgout B, 1998. Mutagenesis of the NS3 protease of dengue virus type 2. *J Virol 72:* 624–632.
42. Murthy HM, Clum S, Padmanabhan R, 1999. Dengue virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects. *J Biol Chem 274:* 5573–5580.
43. Muller DA, Corrie SR, Coffey J, Young PR, Kendall MA, 2012. Surface modified microprojection arrays for the selective extraction of the dengue virus NS1 protein as a marker for disease. *Anal Chem 84:* 3262–3268.
44. Miller S, Sparacio S, Bartenschlager R, 2006. Subcellular localization and membrane topology of the Dengue virus type 2 Non-structural protein 4B. *J Biol Chem 281:* 8854–8863.

45. Modis Y, Ogata S, Clements D, Harrison SC, 2004. Structure of the dengue virus envelope protein after membrane fusion. *Nature 427:* 313–319.
46. Guzman MG, et al., 1995. Partial nucleotide and amino acid sequences of the envelope and the envelope/nonstructural protein-1 gene junction of four dengue-2 virus strains isolated during the 1981 Cuban epidemic. *Am J Trop Med Hyg 52:* 241–246.
47. Rico-Hesse R, et al., 1997. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology 230:* 244–251.
48. Leitmeyer KC, et al., 1999. Dengue virus structural differences that correlate with pathogenesis. *J Virol 73:* 4738–4747.
49. Cologna R, Rico-Hesse R, 2003. American genotype structures decrease dengue virus output from human monocytes and dendritic cells. *J Virol 77*: 3929–3938.
50. Imrie A, et al., 2010. Homology of complete genome sequences for dengue virus type-1, from dengue-fever- and dengue-haemorrhagic-fever-associated epidemics in Hawaii and French Polynesia. *Ann Trop Med Parasitol 104:* 225–235.
51. Samath F. Chinese, Sri Lankan workers mingle at sprawling Hambantota port site. *Financial Times*. Oct. 5, 2008 [online]. Available at: http://www.sundaytimes.lk/081005/FinancialTimes/ft343.html. Accessed April 16, 2014.
52. Mudalige D. Lanka takes over Norochcholai plant. *Features*, Online edition of *Daily News*, Lakehouse Newspapers. Aug. 10, 2011 [online]. Available at: http://archives.dailynews.lk/2011/08/10/fea10.asp. Accessed April 16, 2014.
53. Zheng K, et al., 2009. Molecular characterization of the E gene of dengue virus type 1 isolated in Guangdong province, China, in 2006. *Epidemiol Infect 137:* 73–78.
54. Nicolaisen LE, Desai MM, 2012. Distortions in genealogies due to purifying selection. *Mol Biol Evol 29:* 3589–3600.
55. Jenkins GM, Rambaut A, Pybus OG, Holmes EC, 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol 54:* 156–165.