

Comprehensive Genome Scale Phylogenetic Study Provides New Insights on the Global Expansion of Chikungunya Virus

Rubing Chen,^a Vinita Puri,^b Nadia Fedorova,^b David Lin,^c Kumar L. Hari,^c Ravi Jain,^c Juan David Rodas,^d  Suman R. Das,^b Reed S. Shabman,^b Scott C. Weaver^{a,e}

Institute of Human Infections and Immunology and Department of Pathology, University of Texas Medical Branch, Galveston, Texas, USA^a; J. Craig Venter Institute, Rockville, Maryland, USA^b; CBio Inc., Fremont, California, USA^c; University of Antioquia, Medellín, Colombia^d; Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas, USA^e

ABSTRACT

Since the India and Indian Ocean outbreaks of 2005 and 2006, the global distribution of chikungunya virus (CHIKV) and the locations of epidemics have dramatically shifted. First, the Indian Ocean lineage (IOL) caused sustained epidemics in India and has radiated to many other countries. Second, the Asian lineage has caused frequent outbreaks in the Pacific islands and in 2013 was introduced into the Caribbean, followed by rapid spread to nearly all of the neotropics. Further, CHIKV epidemics, as well as exported cases, have been reported in central Africa after a long period of perceived silence. To understand these changes and to anticipate the future of the virus, the exact distribution, genetic diversity, transmission routes, and future epidemic potential of CHIKV require further assessment. To do so, we conducted the most comprehensive phylogenetic analysis to date, examined CHIKV evolution and transmission, and explored distinct genetic factors associated with the emergence of the East/Central/South African (ECSA) lineage, the IOL, and the Asian lineage. Our results reveal contrasting evolutionary patterns among the lineages, with growing genetic diversities observed in each, and suggest that CHIKV will continue to be a major public health threat with the potential for further emergence and spread.

IMPORTANCE

Chikungunya fever is a reemerging infectious disease that is transmitted by *Aedes* mosquitoes and causes severe health and economic burdens in affected populations. Since the unprecedented Indian Ocean and Indian subcontinent outbreaks of 2005 and 2006, CHIKV has further expanded its geographic range, including to the Americas in 2013. Its evolution and transmission during and following these epidemics, as well as the recent evolution and spread of other lineages, require optimal assessment. Using newly obtained genome sequences, we provide a comprehensive update of the global distribution of CHIKV genetic diversity and analyze factors associated with recent outbreaks. These results provide a solid foundation for future evolutionary studies of CHIKV that can elucidate emergence mechanisms and also may help to predict future epidemics.

Chikungunya virus (CHIKV) (family *Togaviridae*, genus *Alphavirus*) is a mosquito-borne virus that causes a fever-rash-arthritis syndrome in humans, with debilitating joint pain lasting up to several years; the disease is referred to as chikungunya fever (CHIKF) (1). Along with dengue, Zika, and yellow fever viruses, CHIKV is transmitted in human populations by two broadly distributed mosquito vectors, *Aedes aegypti* and *Aedes albopictus* (2–5). Recently, CHIKV established endemic transmission and outbreaks in Africa, Asia, Europe, the Americas, and the Pacific archipelago (6). The rapid expansion of CHIKF, together with its severe morbidity and economic burden, make CHIKV one of the most important arthropod-borne viruses (arboviruses) and a major global health threat.

After suspected outbreaks in Asia and the Americas during the 18th and 19th centuries (7), CHIKV was first isolated during a 1952–1953 epidemic in present-day Tanzania (8, 9). Subsequently, numerous sporadic CHIKF outbreaks have been noted in Africa, with larger epidemics reported in Thailand in the late 1950s and early 1960s and in India from the early 1960s to the early 1970s (10). Whereas the Indian strains disappeared for unknown reasons, the Thailand strains have continued to circulate in Southeast Asia. Phylogenetic studies have revealed that these early CHIKV strains form three major, geographically distinct lineages: the West African enzootic, East/Central/South African (ECSA)

enzootic, and Asian endemic/epidemic lineages (11). The enzootic lineages involve sylvatic transmission by arboreal *Aedes* mosquitoes among nonhuman primates with occasional spillover to humans, while the endemic/epidemic lineages are transmitted human to human by *A. aegypti* and, recently, *A. albopictus* mosquitoes in peridomestic settings (10, 12). Bayesian coalescent analysis has estimated that the Asian and ECSA lineages diverged about 100 years ago, although it is unclear exactly when the Asian lineage was introduced from Africa (13).

The recent spread of CHIKV, particularly two major reemergence events with explosive and extensive spread through long-range international transmission, has brought CHIKV to the forefront of global health attention (14). The first recent emergence

Received 15 June 2016 Accepted 1 September 2016

Accepted manuscript posted online 21 September 2016

Citation Chen R, Puri V, Fedorova N, Lin D, Hari KL, Jain R, Rodas JD, Das SR, Shabman RS, Weaver SC. 2016. Comprehensive genome scale phylogenetic study provides new insights on the global expansion of chikungunya virus. *J Virol* 90:10600–10611. doi:10.1128/JVI.01166-16.

Editor: A. García-Sastre, Icahn School of Medicine at Mount Sinai

Address correspondence to Scott C. Weaver, sweaver@utmb.edu.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

TABLE 1 Strains with full genomes sequenced in this study

Strain	Country	Sample date	Host	Passage history ^a	GenBank accession no.	Sequencing facility
37937	Senegal	1983	<i>A. furcifer</i>	AP-61 1, Vero 2	KX262986	JCVI
37950	Senegal	10 July 1983	<i>A. furcifer</i>	AP-61 1, Vero 1	KX262995	JCVI
6307-88	Thailand	1988	<i>Homo sapiens</i>	LLC-MK2 2, C6/36 1, Vero 1	KX262988	JCVI
667	Cameroon	2006	<i>H. sapiens</i>	Vero 3	KX262996	JCVI
BS 285 (C2)	Malaysia	2009	<i>H. sapiens</i>	C6/36 1	KX262997	JCVI
H 20235-STMARTIN-2013	St. Martin	2013	<i>H. sapiens</i>	? 2, Vero 2	KX262991	JCVI
YO 111213	Guadeloupe	5 January 2014	<i>H. sapiens</i>	C6/36 2, Vero 1	KX262992	JCVI
YO 123223	French Guiana	21 January 2014	<i>H. sapiens</i>	C6/36 2, Vero 1	KX362994	JCVI
Bianchi	Italy	2007	<i>H. sapiens</i>	? 1, Vero 1	KX262989	JCVI
Venturini	Italy	2007	<i>H. sapiens</i>	? 1, Vero 1	KX262993	JCVI
241	Colombia	18 December 2014	<i>H. sapiens</i>	Vero 1	KU365372	UTMB
246	Colombia	December 2014	<i>H. sapiens</i>	Vero 1	KU365373	UTMB

^a AP-61, Vero, LLC-MK2, or C6/36 indicates passage in AP-61, Vero, LLC-MK2, or C6/36 cells, respectively; the following number indicates the number of passages. A question mark indicates an unknown cell line.

involving the ECSA lineage was first recognized with an outbreak in coastal Kenya in 2004, followed by spread to Indian Ocean islands, such as Comoros and La Reunion, causing tens to hundreds of thousands of infections on each island (15). In 2006, this newly emerged epidemic Indian Ocean lineage (IOL) strain spread from East Africa to India, leading to over 1 million cases in the first year alone (16). From India, the IOL spread rapidly to initiate transmission in Southeast Asia, Italy, and France, with exported cases reaching most parts of the world (17, 18).

The second recent CHIKV reemergence involved the endemic Asian lineage. After several decades with few reports of outbreaks or cases in Asia, strains of this lineage caused a series of outbreaks in Malaysia (2006) and Pacific islands, such as New Caledonia (2011); Papua New Guinea (2012); Yap in the Federated States of Micronesia (2013); Tonga, Samoa, American Samoa, Tokelau, and French Polynesia (2014); and Kiribati and the Cook Islands (2015) (5). Then, an Asian lineage strain was detected in the Caribbean island of St. Martin in late 2013 (19), followed by a large-scale epidemic with spread to 26 islands and 14 mainland American countries within 1 year, resulting in more than 1 million suspected cases. As of August 2016, approximately 2 million cases were suspected in 45 countries and territories, according to the Pan American Health Organization. In addition, circulation of a newly imported ECSA strain was also reported for the first time in the Americas, in Brazil in 2014 (20). The virus was imported by a traveler from Angola, but its subsequent distribution in Brazil or elsewhere in the Americas has not been reported.

With these dramatic changes in the global distribution of CHIKV and the emergence of major epidemics, patterns of regional CHIKV circulation, including those of different lineages, paths of spread, evolution, and adaptation, as well as future epidemic risks, require comprehensive assessment. Critical questions include the following. (i) What is the genetic diversity of CHIKV in Africa? Where are the enzootic lineages presently distributed, and how likely is another exportation and global emergence of an enzootic lineage from Africa? (ii) Since the Indian Ocean and Asian outbreaks, where and by what paths has this IOL continued to spread? Furthermore, previous studies have identified convergent selection of a substitution in the E1 envelope glycoprotein (E1-A226V) in multiple clades of the IOL, followed by several E2 envelope glycoprotein substitutions that further adapted some IOL strains to a novel mosquito vector, *A. albopictus*, contributing

to the scale and range of Asian epidemics (21). However, it is unclear what role IOL strains without these adaptive mutations played in the Indian subcontinent and subsequent Southeast Asian outbreaks and to what extent they continue to circulate. (iii) Despite numerous reports of CHIKV outbreaks in the Western Hemisphere associated with the Asian lineage (19, 20, 22–25), their exact transmission route, as well as their phylogenetic relationships with strains from the Pacific islands, have not been fully resolved. It is also unclear whether the emergence of the Asian lineage in the Western Hemisphere was purely a stochastic event or was facilitated by adaptive CHIKV evolution for more efficient urban transmission. Finally, examination of the interplay between Asian and IOL epidemic lineages in Southeast Asia could help us to predict the future distribution of CHIKV genetic diversity and the outcome when multiple lineages cocirculate and potentially compete sympatrically.

To address these questions and provide a timely update on global CHIKV activity, we conducted a comprehensive phylogenetic analysis examining the distinct genetic factors associated with the evolution and emergence of the ECSA and Asian lineages and the IOL. Our results elucidated dramatic movements and previously unsampled genetic diversity, indicating that CHIKV will continue to be a major reemerging global public health threat.

MATERIALS AND METHODS

Virus samples and Illumina sequencing. Twelve CHIKV strains sequenced in this study, listed in Table 1, were obtained from the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) at the University of Texas Medical Branch (UTMB), Galveston, TX. RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA) according to the manufacturers' protocols. Illumina sequencing was performed at the J. Craig Venter Institute (JCVI) or the Genomic Core Facility at UTMB.

At JCVI, sequence-independent single-primer amplification (SISPA) was performed on purified RNA as previously described (26, 27). Bar-coded SISPA products were pooled, and adaptors were ligated following the manufacturers' instructions (Illumina, San Diego, CA). Libraries were sequenced on the Illumina MiSeq (2- by 300-bp paired end [PE]). A combination of JCVI- and CLCbio-developed tools were used to process read data from all next-generation sequencing (NGS) runs (26, 28).

At UTMB, libraries were prepared using the TruSeq RNA sample preparation kit, and Illumina sequencing was performed under the recommended conditions (Illumina). Cluster formation of the library DNA templates was performed using the TruSeq PE Cluster kit v3 and the

TABLE 2 Caribbean strains with 3' UTRs sequenced in this study

Strain	Country	Sample date	Host	Passage history	GenBank accession no.	Duplication
H 20235-STMARTIN-2013	St. Martin	November 2013	<i>Homo sapiens</i>	P2J3, Vero 2	KX262991	Yes
YO 111213	Guadeloupe	January 2014	<i>H. sapiens</i>	C6/36 no. 2, Vero 1	KX262992	Yes
YO 123223	French Guiana	January 2014	<i>H. sapiens</i>	C6/36 no. 2, Vero 1	KX262994	Yes
CH-0008	Mexico	October 2014	<i>H. sapiens</i>	Patient serum	KT327163	Yes
CH-0054	Mexico	October 2014	<i>H. sapiens</i>	Patient serum	KX772762	Yes
CH-0072	Mexico	November 2014	<i>H. sapiens</i>	Patient serum	KT327165	Yes
LI-0031	Mexico	November 2014	<i>H. sapiens</i>	Patient serum	KT327166	Yes
TA-0006	Mexico	November 2014	<i>H. sapiens</i>	Patient serum	KT327167	Yes
N304	Dominican Republic	July 2014	<i>H. sapiens</i>	Patient serum	KX772760	Yes
N456	Dominican Republic	July 2014	<i>H. sapiens</i>	Patient serum	KX772761	Yes
VE 56/14	Trinidad and Tobago	September 2014	<i>H. sapiens</i>	Patient serum	KX772763	Yes
VE 57/2	Trinidad and Tobago	September 2014	<i>H. sapiens</i>	Patient serum	KX772764	Yes
R99659	British Virgin Islands	January 2014	<i>H. sapiens</i>	Vero 2	KX713902	Yes

Illumina cBot workstation under the recommended conditions. Paired-end 50-base sequencing by synthesis was performed using the TruSeq SBS kit v3 on an Illumina HiSeq 1500, following protocols defined by the manufacturer. Sequence analysis and assembly were completed at CBio Inc. (Fremont, CA). Briefly, sequences were first subjected to a BLAST analysis to determine the numbers of viral reads and contaminating reads from hosts or other sources. Next, because virus stocks were prepared on Vero cells, the contaminating reads were filtered using the African green monkey (*Chlorocebus sabaeus*) genome as a template. The remaining Fastq files were further processed using Trimmomatic (72) to remove low-quality sequences. Finally, assembly was performed using iMetAMOS (<http://www.cbcb.umd.edu/software/imetamos>) following the standard parameters.

Sequencing of the 3' UTR. To confirm the 3' untranslated genome region (UTR) structure in the Caribbean/Latin American (referred to below as American) outbreak lineage strains, we sequenced the 3' UTRs of 10 additional samples (Table 2) by the Sanger method. CHIKV RNAs used in this study were directly purified from human serum or from the supernatants of cell cultures after up to two passages in Vero cells. The 3' UTR was amplified by the primer pair Ch21 + 10000 (5'-CACGTAACAGTGATCCCG-3') and Cons-20mer (5'-TTTTTTTGAATATTAATAAAACA AAATA-3') using the Titan One Tube RT-PCR system (Roche, Mannheim, Germany). The amplicons were subsequently gel purified and sequenced using the BigDye Terminator v3.1 cycle-sequencing kit (Applied Biosystems, Foster City, CA) and an ABI Prism model 3100 genetic analyzer (Applied Biosystems). Then, the sequences were edited and assembled in Sequencher v4.9 (Gene Codes, Ann Arbor, MI) and deposited in the GenBank database (Table 2).

Phylogenetic analysis. (i) Data set construction. First, all available CHIKV complete genomic sequences (excluding vaccine and cloning vector strains) available as of February 2016 were downloaded from the GenBank library, aligned using MUSCLE (29), and manually adjusted in Se-AI (available at <http://tree.bio.ed.ac.uk/software/seal/>) according to amino acid sequence alignments to preserve codon homology. Due to the ambiguous alignments of the 3' UTR and the rapid evolution of this genome region, only the open reading frames (ORFs) comprising an alignment of 11,097 nucleotides (nt) were used for phylogenetic tree reconstruction. Second, sequences potentially resulting from artifacts were examined and removed based on a maximum-likelihood (ML) guide tree reconstructed using an initial neighbor-joining tree and a nearest-neighbor interchange (NNI) branch-swapping method implemented in PAUP* (30), based on a GTR+I+ Γ nucleotide substitution model selected by ModelTest (31, 32). These sequences included (i) India/MH4/2000 (nearly identical to Uganda/UgAg4155/1982, suggesting laboratory contamination), (ii) India/STMWG01/2011 (similar to Senegal/A301/1963, also suggesting contamination), and (iii) India/STMWG02/2011 (potentially an assembly error that probably caused *in silico* recombination be-

tween an IOL strain and Senegal/A301/1963). Finally, a few sequences with missing sample years were omitted from the analysis. This resulted in a final data set of 309 sequences. For BEAST analysis, which involves substitution rate estimations, the highly passaged Angola/M2022/1962 strain, which includes an abnormally long terminal branch, probably representing passage-adaptive or random mutations, was removed.

(ii) Phylogenetic reconstruction. A maximum-likelihood tree was first constructed for the complete data set using the methods described above. Due to the preponderant number of sequences in ECSA/IOL (180) and Asian (116) lineages over West African (13) lineages and the lack of new sequences from the last group since our previous 2010 analysis (13), we performed independent analyses of ECSA/IOL and Asian lineages to better understand and compare their evolutionary dynamics using the Bayesian Markov chain Monte Carlo (MCMC) method available in the BEAST v1.8.2 package (18). To exclude the effect of rate variation along branches, which could cause inaccurate estimation of divergence times, we further divided the ECSA lineage into ECSA/African (21 sequences; strains other than IOL, except the two Kenya 2004 strains) and IOL (161 sequences, with two Kenya 2004 strains) data sets and the Asian lineage into Asian/Old (17 sequences; strains before 1996) and Asian/Reemergence (99 sequences; clades I and II, after 2006) data sets. First, to determine the extent of the temporal structure of the sequences and consequently the reliability of our estimates of substitution rates and times to the most recent common ancestors (tMRCAs), we performed a regression analysis of tree root-to-tip genetic distances against sampling dates using the program Path-O-Gen (<http://tree.bio.ed.ac.uk/software/pathogen/>) [17]. This analysis was performed for each lineage based on the corresponding ML tree generated using the methods described above. A strong temporal structural ($R^2 > 0.73$) was observed in each data set. We did not estimate the evolutionary rate for the American lineage due to its lack of temporal structure ($R^2 = 0.15$), reflecting the short period of circulation in the Western Hemisphere.

All the CHIKV strains were dated according to the year and month (if known) of their collection; strains with only a collection year were assigned the median-year value of July. For each data set, BEAST analysis was performed based on either a relaxed molecular clock (uncorrelated lognormal) or a random local clock (33), with the substitution model recommended by jModelTest (34) (GTR+I+ Γ for ECSA/IOL and GTR+ Γ for the Asian lineage). Two demographic models, Bayesian Skyline and Bayesian Skyride, were employed under each clock model, and the results were compared. In each case, MCMC chains were run for a sufficient time to achieve convergence (accessed using the Tracer program [<http://tree.bio.ed.ac.uk/software/tracer/>]). Statistical uncertainty in parameter estimates was reflected as 95% highest probability density (HPD) values. The maximum clade credibility (MCC) tree across all of the plausible trees generated by BEAST was then computed using the TreeAnnotator program available in the BEAST package, with the first 10% of trees

removed as burn in. To illustrate the geographical locations of sample strains and potential transmission relationships, we colored the external branches according to the country/region of origin for each CHIKV strain, with the colors of the internal branches estimated according to the parsimony principle.

Accession number(s). The newly generated full-genome sequences analyzed here have been deposited in GenBank under accession numbers [KU365372](#), [KU365373](#), and [KX262986](#) to [KX262997](#). Newly generated 3' UTR sequences were deposited in GenBank under accession numbers [KX713902](#) and [KX772360](#) to [KX772364](#) or updated based on previous sequences under GenBank accession numbers [KT327163](#) and [KT327165](#) to [KT327167](#).

RESULTS

Evolutionary rates of different CHIKV lineages. To compare the evolutionary patterns of ECSA and Asian lineages, including their different evolutionary periods, and to accurately estimate the tMRCA and divergence time along their evolution history, we divided ECSA into African and IOL data sets, as illustrated in [Fig. 1A](#). The Asian lineage was divided into old (before 1996) and reemerging (after 2006) data sets ([Fig. 1A](#)). Phylogenetic estimations, including the MCC tree, substitution rates, and tMRCA, were performed based on the complete and subdivided data sets for each lineage, with the results summarized in [Table 3](#). Overall, uncorrelated relaxed clock (lognormal) and random local clock models estimated similar evolutionary rates in the range of 1.80×10^{-4} to 8.46×10^{-4} substitutions/site/year. Significant rate variation was observed between African (1.80×10^{-4} to 2.77×10^{-4} substitutions/site/year) and IOL (5.81×10^{-4} to 8.46×10^{-4} substitutions/site/year) subsets in the ECSA lineage, with the latter ~ 3 times higher than the former. In contrast, the differences in evolutionary rates between the old (1958 to 1995) and reemerged (2006 to 2015) subsets in the Asian lineage were smaller: 2.95×10^{-4} to 5.46×10^{-4} substitutions/site/year for old and 3.87×10^{-4} to 7.38×10^{-4} substitutions/site/year for reemerged, although the latter had a larger 95% HPD value than the former.

Correspondingly, the large rate variation between the African and IOL subsets in the ECSA lineage made estimation of the tMRCA and divergence time less reliable using the total data set under the uncorrelated relaxed clock (lognormal) model, where evolutionary rates along each branch are assumed to be clustered around a mean (35). The most recent common ancestor of the ECSA lineage was dated to 1932 to 1951 or 1948 to 1953 based on the total data set according to the Bayesian Skyline or Skyride demographic model, respectively. However, it was dated to 1901 to 1932 or 1904 to 1932 based on the African subset only. This result not only was ~ 30 years earlier than that estimated with the total data set, but also showed no significant difference between the two demographic models. The result indicates that the sylvatic ECSA lineage in Africa and the epidemic IOL have significantly different evolutionary patterns, as discussed in our previous study (13). In contrast, the tMRCA of the American lineage estimated by the reemerged (2012.81 to 2013.74, where numbers after the decimal point represent a fraction of the year; e.g., 81 is in October) and total (2011.80 to 2013.72) Asian data sets were similar. However, the discrepancy in time scale estimation in the ECSA lineage between different data sets was resolved by the random local clock model, which considers rate variation among lineages. Specifically, it provided not only nearly identical estimates of the tMRCA of the IOL based on the total (2003.55 to 2004.38) and IOL (2003.31 to 2004.47) data sets, but also similar estimates of

that of the ECSA lineage based on the total (1900 to 1924) and African (1901 to 1932) data sets. However, due to the complexity of the model, convergence was not achieved when the Skyride demographic model was implemented for the total ECSA data set.

Surprisingly, a lower substitution rate was observed in the re-emerging Asian lineage (3.87×10^{-4} to 7.38×10^{-4} substitutions/site/year) than in the IOL (5.81×10^{-4} to 8.46×10^{-4} substitutions/site/year) despite the fact that both caused urban epidemics. The different evolution rates of the IOL versus the American lineage may reflect their different evolutionary mechanisms (such as adaptive evolution versus population bottlenecks) and/or transmission patterns, as well as amounts of replication; these hypotheses deserve further study.

Evolution of the Asian CHIKV lineage. Overall, the Asian CHIKV lineage in the Eastern Hemisphere revealed a temporal/spatial pattern of evolution with infrequent international transmission. Based on our phylogenetic tree ([Fig. 1B](#)), after the initial introduction during or before the 1950s, the lineage circulated in South and Southeast Asia as geographically segregated sublineages. The Indian sublineage has not been sampled since 1973, presumably reflecting its extinction. Although the Southeast Asian sublineage has not been sampled in Thailand since 1995, it was introduced to Indonesia in the early 1980s and was presumably circulating endemically in the region with occasionally reported cases, such as those in Malaysia in 1998 (36). Since 2006, two deeply divergent clades of the Asian lineage have circulated in Southeast Asia and the Pacific islands, both originating from a 1980s Indonesian strain or a close relative. Clade I was further subdivided into two subclades: one contained strains from 2006–2007 Malaysian outbreaks, and another was sampled from New Caledonia in 2011 and Indonesia in 2013. Clade II contains 2006 Indonesian, 2012 China, 2013 Philippines, 2013 Micronesia, and 2014 American Samoa CHIKV strains. These data suggest that transmission of the Asian lineage in the Eastern Hemisphere has shifted to Indonesia, the South Pacific islands, and probably the Philippines. However, other regions of endemicity may also remain without sampling.

The American CHIKV lineage apparently originated from Asian clade II, with the MRCA estimated between November 2011 and October 2013 ([Table 3](#)). Regardless, the ancestral strain of the American outbreaks diverged from those South Pacific strains a few years before the 2013 Caribbean introduction without detection.

Notably, all Asian lineage CHIKV strains harbor the E1-A98T substitution, which constrains the penetration of the E1-A226V substitution that mediates adaptation to the novel mosquito vector *A. albopictus* in some IOL strains (37). This finding, coupled with a lack in any American CHIKV strain of E1 and E2 *A. albopictus*-adaptive mutations seen in IOL strains (21), as well as a lack of field data implicating *A. albopictus* in transmission in the Americas, corroborates the prediction (38) that this epistatic constraint on the American lineage will restrict it mainly to *A. aegypti* transmission.

CHIKV transmission in the Americas. Due to the rapid spread and explosive transmission of the American CHIKV outbreaks, accompanied by insufficient informative (synapomorphic) mutations, the evolutionary history within the American lineage was not fully resolved. This was reflected by the low posterior support (< 90) of most clades and clusters, indicated by the lack of node bars in [Fig. 1B](#). Nevertheless, multiple American clades were observed, with strains from Martinique and Guadeloupe appearing in the basal position, suggesting an initial Carib-

TABLE 3 Comparison of nucleotide substitution rates and tMRCA among different CHIKV lineages based on different data sets and demographic models

Data set and demographic model	Prior ^a	Substitution rate			tMRCA (yr)											
		Mean	95% HPD		Asian lineage			American lineage			ECSA lineage			IOL		
			Lower	Higher	Mean	Lower	Higher	Mean	Lower	Higher	Mean	Lower	Higher	Mean	Lower	Higher
Asian																
Old.Skyride	URC	4.17E-04	3.14E-04	5.23E-04	1953.76	1948.72	1958.13									
Old.Skyline	URC	4.30E-04	3.21E-04	5.46E-04	1954.28	1948.56	1958.50									
Old.Skyride	RLC	3.68E-04	2.98E-04	4.39E-04	1952.14	1948.60	1955.37									
Old.Skyline	RLC	3.70E-04	2.95E-04	4.45E-04	1952.20	1948.46	1955.87									
Reemergence.Skyride	URC	5.88E-04	4.53E-04	7.38E-04				2013.31	2012.81	2013.68						
Reemergence.Skyline	URC	5.00E-04	3.87E-04	6.21E-04				2013.49	2013.18	2013.74						
Reemergence.Skyride	RLC	5.44E-04	4.35E-04	6.50E-04				2013.33	2012.94	2013.64						
Reemergence.Skyline	RLC	5.15E-04	4.23E-04	6.13E-04				2013.51	2013.24	2013.73						
Total.skyride	URC	3.86E-04	3.20E-04	4.51E-04	1955.04	1952.02	1957.77	2012.53	2011.80	2013.19						
Total.skyline	URC	4.41E-04	3.82E-04	4.98E-04	1954.17	1949.91	1957.70	2013.46	2013.14	2013.72						
Total.skyride	RLC	4.11E-04	3.61E-04	4.60E-04	1954.40	1952.38	1956.34	2012.80	2012.19	2013.38						
Total.skyline	RLC	4.38E-04	3.95E-04	4.86E-04	1954.26	1952.12	1956.17	2013.43	2013.11	2013.70						
ECSA																
African.skyride	URC	2.32E-04	1.86E-04	2.77E-04							1918.63	1903.62	1932.22			
African.skyline	URC	2.27E-04	1.80E-04	2.71E-04							1916.51	1900.90	1932.10			
African.skyride	RLC	2.19E-04	1.86E-04	2.52E-04							1913.76	1902.69	1924.43			
African.skyline	RLC	2.16E-04	1.82E-04	2.50E-04							1912.58	1900.34	1923.57			
IOL.skyride	URC	7.51E-04	6.67E-04	8.46E-04										2004.20	2003.90	2004.47
IOL.skyline	URC	7.37E-04	6.58E-04	8.22E-04										2003.93	2003.31	2004.41
IOL.skyride	RLC	6.53E-04	5.87E-04	7.20E-04										2004.10	2003.81	2004.38
IOL.skyline	RLC	6.50E-04	5.81E-04	7.19E-04										2003.96	2003.55	2004.31
Total.skyride	URC	4.82E-04	4.34E-04	5.24E-04							1951.23	1948.43	1953.08	2003.45	2002.54	2004.16
Total.skyline	URC	4.53E-04	4.02E-04	5.01E-04							1942.01	1931.52	1951.90	2003.49	2002.61	2004.22
Total.skyride ^b	RLC															
Total.skyline	RLC	3.30E-04	2.95E-04	3.66E-04							1919.21	1909.38	1928.03	2002.31	2001.01	2003.61

^a URC, uncorrelated relaxed clock (lognormal); RLC, random local clock.

^b Bayesian MCMC runs could not reach convergence; therefore, no reliable result was obtained.

bean source of the outbreak. It is unclear, however, whether this basal position relative to St. Martin is correct due to the paucity of strains from the latter location, where CHIKV was first reported in the Americas. From there, American lineage strains spread to other Caribbean islands and then to Central, North, and South America. This multiclade structure indicated that initial transmission was multidirectional. Each of these American clades included strains from multiple countries or territories, and several countries had cocirculating strains from different clades, such as Mexico, Trinidad, Brazil, and Colombia.

In summary, CHIKV strains in the Western Hemisphere did not form geographically isolated clades but rather represented an intertwined transmission network among different countries and regions. Surprisingly, a strain from French Polynesia collected in January 2015 also fell within the American lineage, with a Virgin Islands October 2014 strain as its closest neighbor. This indicates that CHIKV transmission between the South Pacific and Caribbean islands may occur frequently and bidirectionally, suggesting that the initial Caribbean outbreak strain could have originated from a South Pacific island.

Unique genetic structures of the American CHIKV lineage.

Two distinct genetic characters were observed in the American lineage. As previously reported, a deletion in nsP3 codons 379 to 382 was observed in all American lineage strains, along with other Asian clade II strains (Fig. 1B). Surprisingly, a larger, overlapping deletion in nsP3 codons 376 to 382 was observed in another clade currently circulating in Indonesia and New Caledonia, as well as in some 2006-2007 Malaysian strains. It is not clear if these two deletions share an origin (with the shorter 379-to-382 deletion presumably preceding a subsequent 376-to-378 deletion) and why

only some of the Malaysian 2006-2007 strains have the deletion. Based on the parsimony principle, recombination between a Malaysian CHIKV strain and the nsP3 376-to-382 deletion and another cocirculating strain without the deletion could have generated the Malaysian 2006-2007 strains lacking the deletion.

In addition to the nsP3 deletion, an insertion of 177 nt (39), presumably resulting from duplication of the immediately adjacent region, was observed in the 3' UTRs of all American CHIKV strains (Fig. 1B). Due to the assembly of Illumina reads using a non-Caribbean reference strain, this insertion was overlooked in the early Caribbean sequences deposited in GenBank. Independent reports of an elongated 3' UTR in a few Brazilian and Caribbean strains prompted us to confirm by Sanger sequencing that all American 3' UTRs have the same insertion. As indicated in Fig. 1B, these strains were distributed in all American lineage clades, as well as at the basal position. This finding, together with the absence of the insertion in Asian lineage strains from the Eastern Hemisphere, indicates that the insertion most likely occurred as a founder effect during the introduction from the Eastern Hemisphere, presumably by an infected traveler or mosquito.

ECSA lineage in Africa and new exportations. The ECSA lineage was named based on its initial detection mainly in East, Central, and South Africa. Despite the recent 2004 epidemic in East Africa (coastal Kenya), giving rise to the Indian Ocean and Indian subcontinent epidemics, little information is known about (i) where ECSA strains are currently circulating enzootically and/or endemically; (ii) whether the Kenya 2004 strains spread to other African regions; and (iii) if the IOL has spread in Africa and potentially replaced the historical ECSA strains there, given its dramatic spread in Asia, Oceania,

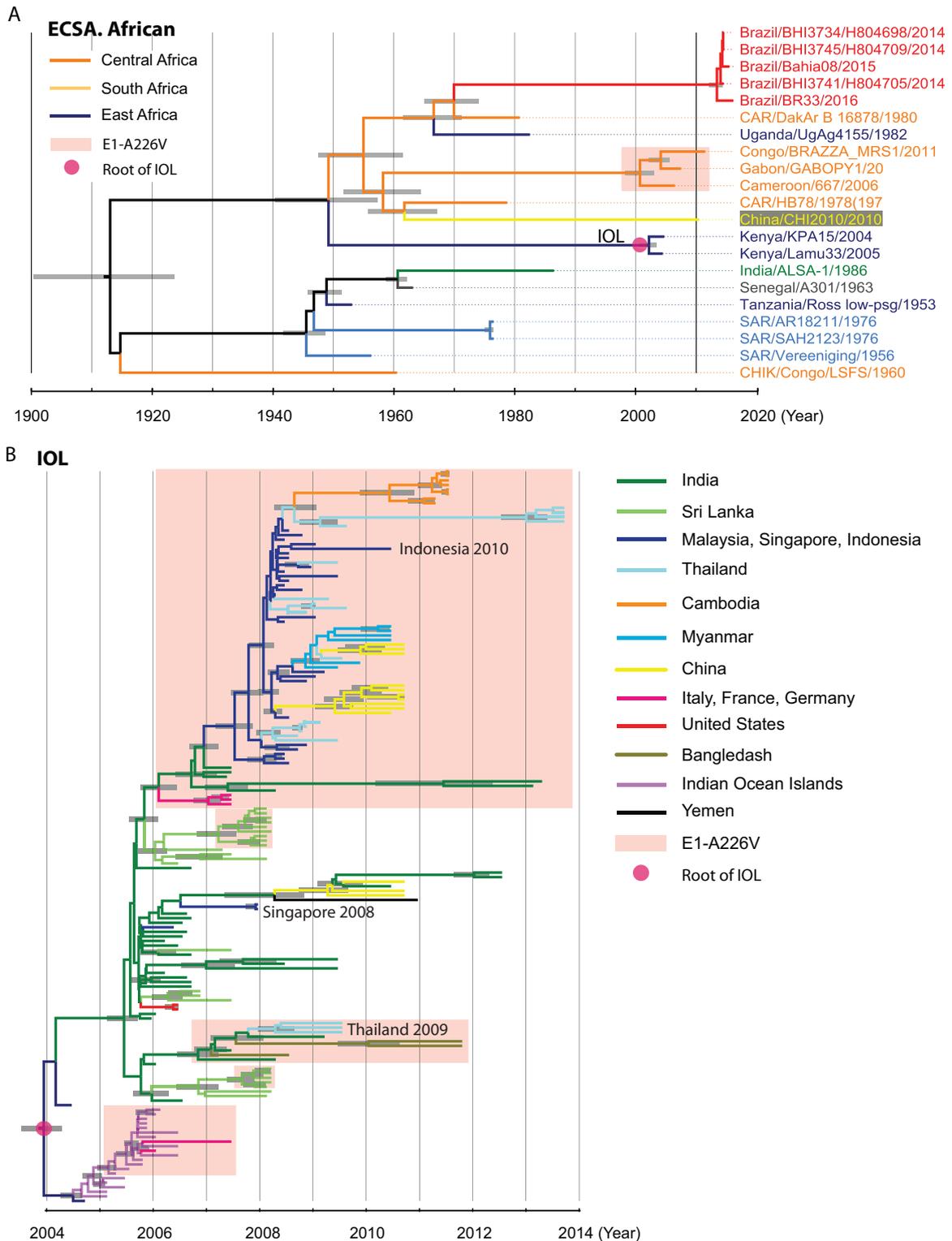


FIG 2 Evolution of the CHIKV ECSA lineage. (A) MCC tree of the CHIKV ECSA lineage in Africa, excluding the IOL and including all African and a few exported strains, other than the IOL, based on random local clocks and the Skyline population model. (B) MCC tree of the CHIKV IOL based on random local clocks and the Skyline population model. The branch lengths are scaled to sampling and divergence times, and the branches are color coded for sample location according to the legend. The gray horizontal node bars representing 95% HPD values of the node height are shown only for those with posterior possibilities of ≥ 90 .

and Europe. The availability of newly derived sequences of ECSA origin allowed us to address these questions.

Based on our phylogenetic tree (Fig. 2A), ECSA strains have not been collected from South Africa and analyzed since 1976. Our results also provide no evidence for further spread of the Kenya 2004 strain in Africa or of the reintroduction of IOL strains into Africa. Rather, newly identified ECSA strains/sublineages were observed in Central Africa (Cameroon, Gabon, and Congo), China, and Brazil. All of these strains originated from a Central African clade, and the lack of close relatives of the strains indicates cryptic endemic/enzootic transmission in the region. Strikingly, instead of forming a single clade, these newly sampled strains were all deeply rooted in the 1978–1982 Central African clade as three independent branches. This finding, together with the deep divergence between the Central African clade and the IOL-ancestral Kenya 2004 strain, suggests that East and Central Africa harbor rich, largely unsampled CHIKV genetic diversity, with deeply divergent strains cocirculating in these regions. The repeated exportation of ECSA strains to Asia (13), Indian Ocean islands (15), Europe (40–42), and South America (20), as well as ever-increasing intercontinental transportation, suggests that the ECSA lineage will seed additional epidemics in the near future. Importantly, the Cameroon (2006), Gabon (2007), and Congo (2011) strains derived from the Central African clade of the ECSA lineage all contain the E1-A226V *A. albopictus*-adaptive substitution, suggesting a high risk of urban transmission in locations where this vector is more abundant than *A. aegypti*, such as southern Europe (43).

Evolution and expansion of the IOL. Since the unprecedented 2005–2006 epidemics in India and the Indian Ocean islands, IOL strains have rapidly spread to Southeast Asia and have been exported to many other countries in Asia, Europe, the Middle East, Oceania, and the Americas. The current circulation patterns in these regions require further investigation. Consistent with our previous study, the IOL consists of two independent sublineages: the Indian Ocean islands sublineage and the India sublineage that emerged independently from coastal Kenya and later spread to Southeast Asia, Italy, and France (Fig. 2B). In the Indian Ocean islands, after the initial major outbreak in 2005 and 2006, there were no further reports of CHIKV outbreaks until the CHIKV reemergence in La Reunion Island in 2010 (no genome sequence is available) (44). However, the case exported to Germany in 2007 indicates that a low level of endemic transmission persisted. On the other hand, the IOL Indian sublineage continued endemic/epidemic circulation in India and radiated to cause additional outbreaks in surrounding countries, such as Sri Lanka, Bangladesh, Yemen, and China. Importantly, IOL strains have established persistent transmission in Southeast Asia. During 2008 and 2009, despite at least three exportations from India to Southeast Asian countries, such as Singapore, Thailand, and Malaysia (Fig. 2B), the major IOL sublineage that caused simultaneous outbreaks in Malaysia, Thailand, and Singapore appears to have established persistent transmission in Thailand, as suggested by inclusion of the 2013 Thailand strains in the same clade. In addition, this sublineage also spread to Myanmar, China, and Cambodia. Clearly, Southeast Asia remains an endemic reservoir for potential global IOL introductions.

Previous studies demonstrated convergent selection of E1-A226V in both the Indian Ocean islands and four independent clades in India that were identified during subsequent outbreaks. We examined the progress and distribution of this substitution. As shown in Fig. 2B, India contained the most CHIKV genetic diver-

sity, with both E1-226A and E1-226V variants actively circulating, and both have been repeatedly exported to surrounding countries. Similarly, the convergent E1-A226V substitution occurred in two independent clades after introduction of the IOL to Sri Lanka. However, due to the lack of recent sequences, the genetic diversity of CHIKV in South Asia is unclear, i.e., whether either or both clades and/or the E1-226 variants coexist. In contrast, in Southeast Asia, the IOL strains mainly comprise a single monophyletic clade that acquired E1-226V before its introduction from South Asia.

DISCUSSION

The complex phylogenetic structure of the ECSA lineage indicates deep genetic diversity and further emergence risk. The evolutionary patterns and emergence potential of arboviruses are directly related to their transmission cycles. In Africa, CHIKV circulates in a sylvatic cycle among nonhuman primates, mediated by forest-dwelling mosquitoes, such as *Aedes furcifer*, *Aedes africanus*, and *Aedes luteocephalus* (45). Human infections are considered mainly incidental spillovers, although endemic circulation may occur for extended periods. Outside Africa, CHIKV is transmitted among human populations via two anthropophilic mosquitoes, *A. aegypti* and *A. albopictus*, in urban transmission cycles. Not surprisingly, distinct phylogenetic patterns were observed between the sylvatic/enzootic ECSA lineage and the epidemic IOL and American lineage. On one hand, the shallow bush-like trees of the IOL and the American lineage result from founder effects, followed by radiation due to rapid and explosive expansion into a large geographic area, with intertwined branches reflecting frequent genetic exchange among different regions, presumably by travelers. In contrast, the phylogenetic structure of the ECSA strains in Africa suggests deep, longstanding genetic diversity with long regionally defined branches that diverged about 40 to 50 years ago, reflecting limited genetic exchange among geographic regions of Africa.

The paucity of reported endemic/enzootic CHIKV transmission in Africa despite cases exported from distinct origins suggests additional, undetected genetic diversity. For example, despite the close genetic relationship of CHIKV strains that caused autochthonous 2014 transmission in Brazil to an Angola 1962 strain, no outbreaks have been reported from Angola or surrounding regions since 1973. Similarly, the origin of the exported 2010 China ECSA strain remains unknown because closely related strains have not been identified in Africa. Furthermore, the lack of CHIKV strains closely related to those that caused the Kenya 2004 outbreak, which seeded the Indian Ocean and Indian outbreaks, underscores unsampled CHIKV circulation in East Africa.

Nevertheless, the ability of ECSA CHIKV strains to seed outbreaks overseas, as well as occasionally recognized local outbreaks in African cities, such as in Cameroon, Gabon, and Congo, where *A. aegypti* and *A. albopictus* were identified as vectors, indicates that ECSA strains can readily initiate urban transmission. Indeed, the emergence of the E1-A226V mutation in the Cameroon/Gabon/Congo monophyletic clade, as well as the identification of *A. albopictus* as a vector in some African outbreaks (46–48), suggests that this mutation is probably selected by the same mechanism that occurred in the IOL. The frequent detection of outbreaks in Central African cities over a relatively short term (2006 to 2011) also suggests that this ECSA clade has established continuous urban transmission in Africa and may now be evolving independently of enzootic transmission. Not surprisingly, this ECSA clade seeded autochthonous *A. albopictus*-borne transmission in Mont-

pellier, France, from September to October 2014, introduced by a traveler returning from Cameroon (40). The establishment of an urban CHIKV lineage in Africa, particularly one that can be transmitted by both *A. aegypti* and *A. albopictus*, poses a great risk for further global exportation and emergence.

Determinants of the emergence of the Asian CHIKV lineage in the Caribbean. Perhaps the most intriguing question regarding the CHIKV outbreak in the Americas is why the Asian lineage eventually invaded the Americas rather than the IOL, which was introduced years earlier into the Americas by thousands of infected travelers. Did genetic, ecological/geographical, or stochastic factors play more important roles in this establishment?

Genetically, the American lineage CHIKV strains evolved from an ancestor with two unique characteristics. The 3' UTR insertion, which represents an approximate duplication of the previously defined DR(1 + 2) region, has been shown to improve CHIKV fitness for replication in mosquito cells without a significant effect in Vero cells (39). Previous *in vivo* studies with *A. aegypti* mosquitoes and CD-1 infant mice suggested a similar fitness effect of DR(1 + 2) duplication in mosquitoes but a slightly deleterious phenotype in mammals (49). Therefore, to some extent, increased fitness in *A. aegypti* mosquitoes, the only vector directly implicated in CHIKV transmission in the Americas (50), mediated by this duplication may have facilitated transmission and spread from the Caribbean. The selection of longer CHIKV 3' UTRs has been shown to occur mainly during replication in the mosquito host (49), suggesting that this duplication probably arose in the vector. However, the putative fitness advantage was apparently not strong enough for establishment in South Pacific islands, as suggested by the predominance of Asian lineage strains with the shorter 3' UTR there. Therefore, the duplication may have arisen in Oceania but been fixed in the Americas due to a founder effect after introduction either by a traveler or a mosquito. On the other hand, it is unclear whether the Asian lineage nsP3 deletion confers any fitness advantage in any host or vector. The alphaviral nsP3 protein contains two distinct domains: the highly conserved N-terminal macrodomain and the hypervariable domain (HVD), which is devoid of any predicted secondary structure (51). The nsP3 deletions in the Asian lineage fall into the hypervariable domain, which is highly plastic during alphavirus evolution. Due to its rapid evolution, the HVD is hypothesized to be involved in the optimization of replication in diverse host cell types (51). Chimeric viruses containing CHIKV and o'nyong-nyong virus (ONNV) genes showed that, in the CHIKV backbone, the ONNV nsP3 gene confers the ability to infect *Anopheles* mosquitoes, natural vectors of this close relative of CHIKV (52). Another study suggested that deletion of the HVD of Sindbis virus causes it to be defective for mosquito cell infection (53). Attenuated virulence and reduced replication rates of Semliki Forest virus mutants lacking some portion of the nsP3 HVD were also observed in vertebrate cells (54). However, a 34-amino-acid deletion in the nsP3 region in Venezuelan equine encephalitis virus had no detectable effect on replication in vertebrate cells. In summary, the effect of nsP3 appears to be host and virus specific, so the potential fitness effect of the nsP3 deletion in American CHIKV strains requires further study.

An important yet understudied observation is the close phylogenetic relationship between CHIKV strains from the Pacific and those in the Caribbean islands. The Asian lineage strains most closely related to those from the Caribbean are from South Pacific islands and the Philippines. In addition, a 2015 strain from French

Polynesia in the South Pacific was apparently introduced back from the Caribbean, indicating frequent transmission between the two regions. Interestingly, the spread of Zika virus seems have taken the same route: from Pacific islands to the Americas coincident with the introduction of CHIKV to the Caribbean (55).

Despite the frequent international transmission of infectious agents, adjacent regions that share similar ecological systems and vector populations still provide the best chance for the spread of arboviruses. It is the frequent genetic exchange, combined with suitable ecological conditions for efficient transmission, that promotes the possibility of an outbreak after CHIKV introduction. Given that the IOL is currently circulating in Southeast Asia, including Thailand, Cambodia, and Malaysia, with few reports in Indonesia and the Philippines (see below), better surveillance to monitor the progress of IOL in Southeast Asia may help to predict its potential for future expansion, especially for temperate regions inhabited by *A. albopictus* but not *A. aegypti*.

Cocirculation of different lineages and potential outcomes.

Due to the recent CHIKV geographic expansions, lineages or clades with different virulences or vector preferences have come to cocirculate. First, although the E1-A226V mutation has been selected in different IOL strains multiple times and has been rapidly distributed to many countries in South and Southeast Asia, in certain areas of India, such as Andhra Pradesh State, only strains with E1-226A were observed (56, 57). The predominance of E1-226A versus E1-226V in different states parallels the abundance of *A. aegypti* and *A. albopictus* (58), indicating that vector preference plays an important role in IOL distribution.

The cocirculation of the IOL and the Asian lineage in Southeast Asia appears to be more complicated. CHIKF outbreaks in Southeast Asia have been markedly episodic, with long interepidemic intervals. This can be at least partially attributed to durable immunity, as suggested by serosurveys in Thailand (59), the Philippines (60, 61), and Cambodia (62). However, the recent introduction into the region of the IOL has resulted in large-scale epidemics in both Malaysia and Thailand in 2008 and 2009. In Thailand, the Asian CHIKV lineage caused significant outbreaks in the past but has not been reported since 1995. The 2008-2009 Thailand outbreak, caused by IOL strains, led to 46,000 reported cases in 43 out of 75 provinces. In Malaysia, the Asian lineage caused small, regional CHIKF outbreaks in 1998 and in 2006 and 2007 (63). In November 2006, a small outbreak was reportedly caused by an IOL strain imported by a traveler from India (64) (no CHIKV genome sequence is available). However, two independent IOL clades (Fig. 2B) caused national-scale outbreaks in 2008 and 2009 (65, 66). Experimental studies have demonstrated that Asian and IOL strains generate cross-protective immunity (66); therefore, the rapid spread of the IOL in Thailand and Malaysia could be attributed to its higher replication rate than the Asian lineage, as suggested by *in vitro* data (49). However, in Indonesia and the Philippines, despite the detection of IOL strains in 2011 (67) and 2013 (68), large-scale outbreaks during this period were still dominated by the Asian lineage. The factors that caused the different epidemiological patterns of the 2 epidemic CHIKV lineages among Southeast Asian countries require further analysis.

Similar cocirculation of ECSA and Asian lineage CHIKV strains has occurred in Brazil since 2014, when both American and ECSA lineage strains arrived to initiate autochthonous transmission in Oiapoque, state of Amapá (northeast), and Feira de Santana, state of Bahia (southeast), respectively. Modeling studies

(20) suggest the potential for these lineages to overlap in the same geographical region. Recent sequences show that the ECSA lineage continued to circulate in the state of Bahia as of July 2015 and spread south to the state of Pernambuco in March 2016. On the other hand, an American strain caused the first autochthonous transmission in the state of Rio de Janeiro in December 2015 (69). Our phylogenetic trees (Fig. 1B) suggest that the Rio strain is closely related to a Brazilian strain from the state of Pernambuco, although it is unclear if it is a direct descendant. The continued spread and interactions of these strains demand additional monitoring. Of potential importance is the presence in the ECSA Brazilian strain of the E1-211I residue compared to E1-211T found in IOL strains. This mutation has been shown to constrain the ability of CHIKV to infect *A. albopictus* in the presence of E1-226V (70).

Finally, CHIKV has recently been exported to many other countries and caused local transmission in some. For example, the IOL caused outbreaks in Italy in 2007 (41) and France in 2010 (42). Furthermore, ECSA, Asian, and IOL strains have been introduced into China, with the IOL introduced multiple times from different origins (Fig. 1 and 2). Although CHIKV is not known to have established endemic transmission in any of these countries, suitable vector populations and climates suggest the potential for repeated introductions and epidemics.

In summary, the dramatic global expansion of CHIKV since 2004, together with the continuing spread of the mosquito vectors *A. aegypti* and *A. albopictus*, such as in Russia (71), indicates that CHIKV will continue to expand its global distribution and become an increasing threat to public health.

ACKNOWLEDGMENTS

We thank Philippe Lemey and Edward Holmes for their advice and Mathilde Guerbois-Galla for technical help.

This project was funded in part by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under awards AI120942, AI121452, and U19AI110819. Some of the genomes were sequenced with funding provided by DHS S&T through contract no. HSHQDC-13-C-B0009 (Capturing Global Biodiversity of Pathogens by Whole Genome Sequencing).

The content is solely our responsibility and does not necessarily represent the official views of the National Institutes of Health.

FUNDING INFORMATION

This work, including the efforts of David Lin, Kumar Hari, and Ravi Jain, was funded by Department of Homeland Security (HSHQDC-13-C-B0009). This work, including the efforts of Rubing Chen, was funded by Department of Homeland Security (HSHQDC-13-C-B0009). This work, including the efforts of Scott C. Weaver, was funded by Department of Homeland Security (HSHQDC-13-C-B0009). This work, including the efforts of Scott C. Weaver, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (AI120942 and AI121452). This work, including the efforts of Rubing Chen, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (AI121452). This work, including the efforts of Vinita Puri and Nadia Fedorova, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (U19AI110819). This work, including the efforts of Suman R. Das and Reed S. Shabman, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (U19AI110819).

REFERENCES

- Weaver SC, Lecuit M. 2015. Chikungunya virus and the global spread of a mosquito-borne disease. *N Engl J Med* 372:1231–1239. <http://dx.doi.org/10.1056/NEJMra1406035>.
- Weaver SC, Costa F, Garcia-Blanco MA, Ko AI, Ribeiro GS, Saade G, Shi PY, Vasilakis N. 2016. Zika virus: history, emergence, biology, and prospects for control. *Antiviral Res* 130:69–80. <http://dx.doi.org/10.1016/j.antiviral.2016.03.010>.
- Chen R, Vasilakis N. 2011. Dengue—quo tu et quo vadis? *Viruses* 3:1562–1608. <http://dx.doi.org/10.3390/v3091562>.
- Tsetsarkin KA, Chen R, Sherman MB, Weaver SC. 2011. Chikungunya virus: evolution and genetic determinants of emergence. *Curr Opin Virol* 1:310–317. <http://dx.doi.org/10.1016/j.coviro.2011.07.004>.
- Weaver SC, Reisen WK. 2010. Present and future arboviral threats. *Antiviral Res* 85:328–345. <http://dx.doi.org/10.1016/j.antiviral.2009.10.008>.
- Petersen LR, Powers AM. 2016. Chikungunya: epidemiology. *F1000Res* 5:F1000 Faculty Rev-82. <http://dx.doi.org/10.12688/f1000research.7171.1>.
- Halstead SB. 2015. Reappearance of chikungunya, formerly called dengue, in the Americas. *Emerg Infect Dis* 21:557–561. <http://dx.doi.org/10.3201/eid2104.141723>.
- Ross RW. 1956. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. *J Hyg* 54:177–191.
- Lumsden WH. 1955. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53. II. General description and epidemiology. *Trans R Soc Trop Med Hyg* 49:33–57. [http://dx.doi.org/10.1016/0035-9203\(55\)90081-X](http://dx.doi.org/10.1016/0035-9203(55)90081-X).
- Weaver SC, Forrester NL. 2015. Chikungunya: evolutionary history and recent epidemic spread. *Antiviral Res* 120:32–39. <http://dx.doi.org/10.1016/j.antiviral.2015.04.016>.
- Powers AM, Brault AC, Tesh RB, Weaver SC. 2000. Re-emergence of Chikungunya and O'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *J Gen Virol* 81:471–479. <http://dx.doi.org/10.1099/0022-1317-81-2-471>.
- Weaver SC. 2013. Urbanization and geographic expansion of zoonotic arboviral diseases: mechanisms and potential strategies for prevention. *Trends Microbiol* 21:360–363. <http://dx.doi.org/10.1016/j.tim.2013.03.003>.
- Volk SM, Chen R, Tsetsarkin KA, Adams AP, Garcia TI, Sall AA, Nasar F, Schuh AJ, Holmes EC, Higgs S, Maharaj PD, Brault AC, Weaver SC. 2010. Genome-scale phylogenetic analyses of chikungunya virus reveal independent emergences of recent epidemics and various evolutionary rates. *J Virol* 84:6497–6504. <http://dx.doi.org/10.1128/JVI.01603-09>.
- Morens DM, Fauci AS. 2014. Chikungunya at the door: deja vu all over again? *N Engl J Med* 371:885–887. <http://dx.doi.org/10.1056/NEJMp1408509>.
- Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, Lavenir R, Pardigon N, Reynes JM, Pettinelli F, Biscornet L, Diancourt L, Michel S, Duquerroy S, Guignon G, Frenkiel MP, Brehin AC, Cubito N, Despres P, Kunst F, Rey FA, Zeller H, Brisse S. 2006. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med* 3:e263. <http://dx.doi.org/10.1371/journal.pmed.0030263>.
- Arankalle VA, Shrivastava S, Cherian S, Gunjekar RS, Walimbe AM, Jadhav SM, Sudeep AB, Mishra AC. 2007. Genetic divergence of Chikungunya viruses in India (1963–2006) with special reference to the 2005–2006 explosive epidemic. *J Gen Virol* 88:1967–1976. <http://dx.doi.org/10.1099/vir.0.82714-0>.
- Panning M, Grywna K, van Esbroeck M, Emmerich P, Drosten C. 2008. Chikungunya fever in travelers returning to Europe from the Indian Ocean region, 2006. *Emerg Infect Dis* 14:416–422. <http://dx.doi.org/10.3201/eid1403.070906>.
- Lanciotti RS, Kosoy OL, Laven JJ, Panella AJ, Velez JO, Lambert AJ, Campbell GL. 2007. Chikungunya virus in US travelers returning from India, 2006. *Emerg Infect Dis* 13:764–767. <http://dx.doi.org/10.3201/eid1305.070015>.
- Leparc-Goffart I, Nougaiere A, Cassadou S, Prat C, de Lamballerie X. 2014. Chikungunya in the Americas. *Lancet* 383:514. [http://dx.doi.org/10.1016/S0140-6736\(14\)60185-9](http://dx.doi.org/10.1016/S0140-6736(14)60185-9).
- Nunes MR, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, Azevedo RDS, da Silva DE, da Silva EV, da Silva SP, Carvalho VL, Coelho GE, Cruz AC, Rodrigues SG, da Silva Goncalves Vianez JL, Jr, Nunes BT, Cardoso JF, Tesh RB, Hay SI, Pybus OG, da Costa Vasconcelos PF. 2015. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med* 13:102. <http://dx.doi.org/10.1186/s12916-015-0348-x>.
- Tsetsarkin KA, Chen R, Yun R, Rossi SL, Plante KS, Guerbois M, Forrester N, Perng GC, Sreekumar E, Leal G, Huang J, Mukhopad-

- hyay S, Weaver SC. 2014. Multi-peaked adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in *Aedes albopictus* mosquitoes. *Nat Commun* 5:4084. <http://dx.doi.org/10.1038/ncomms5084>.
22. Kautz TF, Diaz-Gonzalez EE, Erasmus JH, Malo-Garcia IR, Langsjoen RM, Patterson EI, Auguste DI, Forrester NL, Sanchez-Casas RM, Hernandez-Avila M, Alpuche-Aranda CM, Weaver SC, Fernandez-Salas I. 2015. Chikungunya virus as cause of febrile illness outbreak, Chiapas, Mexico, 2014. *Emerg Infect Dis* 21:2070–2073. <http://dx.doi.org/10.3201/eid2111.150546>.
 23. Wang C, Saborio S, Gresh L, Eswarappa M, Wu D, Fire A, Parameswaran P, Balmaseda A, Harris E. 2016. Chikungunya virus sequences across the first epidemic in Nicaragua, 2014–2015. *Am J Trop Med Hyg* 94:400–403. <http://dx.doi.org/10.4269/ajtmh.15-0497>.
 24. Sahadeo N, Mohammed H, Allicock OM, Auguste AJ, Widen SG, Badal K, Pulchan K, Foster JE, Weaver SC, Carrington CV. 2015. Molecular characterisation of chikungunya virus infections in Trinidad and comparison of clinical and laboratory features with dengue and other acute febrile cases. *PLoS Negl Trop Dis* 9:e0004199. <http://dx.doi.org/10.1371/journal.pntd.0004199>.
 25. Mattar S, Miranda J, Pinzon H, Tique V, Bolanos A, Aponte J, Arrieta G, Gonzalez M, Barrios K, Contreras H, Alvarez J, Aleman A. 2015. Outbreak of Chikungunya virus in the north Caribbean area of Colombia: clinical presentation and phylogenetic analysis. *J Infect Dev Ctries* 9:1126–1132. <http://dx.doi.org/10.3855/jidc.6670>.
 26. Moser LA C-RL, Puri V, Pauszek SJ, Matthews K, Dilley K, Mullan C, McGraw J, Khayat M, Beeri K, Yee A, Dugan V, Heise MT, Frieman MB, Rodriguez LL, Bernard KA, Wentworth DE, Stockwell MB, Shabman RS. 2016. A universal next generation sequencing protocol to generate non-infectious barcoded cDNA libraries from high containment RNA viruses. *mSystems* 1:e00039-15. <http://dx.doi.org/10.1128/mSystems.00039-15>.
 27. Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, Spiro DJ. 2008. Viral genome sequencing by random priming methods. *BMC Genomics* 9:5. <http://dx.doi.org/10.1186/1471-2164-9-5>.
 28. Stockwell TB, Heberlein-Larson LA, Tan Y, Halpin RA, Fedorova N, Katzell DA, Smole S, Unnasch TR, Kramer LD, Das SR. 2015. First complete genome sequences of two keystone viruses from Florida. *Genome Announc* 3:e01255-15. <http://dx.doi.org/10.1128/genomeA.01255-15>.
 29. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <http://dx.doi.org/10.1186/1471-2105-5-113>.
 30. Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics* Chapter 6:Unit 6.4.
 31. Posada D. 2003. Using MODELTEST and PAUP* to select a model of nucleotide substitution. *Curr Protoc Bioinformatics* Chapter 6:Unit 6.5.
 32. Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818. <http://dx.doi.org/10.1093/bioinformatics/14.9.817>.
 33. Drummond AJ, Suchard MA. 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biol* 8:114. <http://dx.doi.org/10.1186/1741-7007-8-114>.
 34. Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253–1256. <http://dx.doi.org/10.1093/molbev/msn083>.
 35. Ho SY, Duchene S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol Ecol* 23:5947–5965. <http://dx.doi.org/10.1111/mec.12953>.
 36. Hasebe F, Parquet MC, Pandey BD, Mathenge EG, Morita K, Balasubramaniam V, Saat Z, Yusop A, Sinniah M, Natkunam S, Igarashi A. 2002. Combined detection and genotyping of Chikungunya virus by a specific reverse transcription-polymerase chain reaction. *J Med Virol* 67:370–374. <http://dx.doi.org/10.1002/jmv.10085>.
 37. Tssetsarkin KA, Chen R, Leal G, Forrester N, Higgs S, Huang J, Weaver SC. 2011. Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc Natl Acad Sci U S A* 108:7872–7877. <http://dx.doi.org/10.1073/pnas.1018344108>.
 38. Weaver SC. 2014. Arrival of chikungunya virus in the new world: prospects for spread and impact on public health. *PLoS Negl Trop Dis* 8:e2921. <http://dx.doi.org/10.1371/journal.pntd.0002921>.
 39. Stapleford KA, Moratorio G, Henningson R, Chen R, Matheus S, Enfissi A, Weissglas-Volkov D, Isakov O, Blanc H, Mounce BC, Dupont-Rouzeyrol M, Shomron N, Weaver S, Fontes M, Rousset D, Vignuzzi M. 2016. Whole-genome sequencing analysis from the chikungunya virus Caribbean outbreak reveals novel evolutionary genomic elements. *PLoS Negl Trop Dis* 10:e0004402. <http://dx.doi.org/10.1371/journal.pntd.0004402>.
 40. Delisle E, Rousseau C, Broche B, Leparac-Goffart I, L'Ambert G, Cochet A, Prat C, Foulongne V, Ferre JB, Catelinois O, Flusin O, Tchernonog E, Moussion IE, Wiegandt A, Septfons A, Mendy A, Moyano MB, Laporte L, Maurel J, Jourdain F, Reynes J, Paty MC, Golliot F. 2015. Chikungunya outbreak in Montpellier, France, September to October 2014. *Euro Surveill* 20:21108. <http://dx.doi.org/10.2807/1560-7917.ES2015.20.17.21108>.
 41. Angelini R, Finarelli AC, Angelini P, Po C, Petropulacos K, Macini P, Fiorentini C, Fortuna C, Venturi G, Romi R, Majori G, Nicoletti L, Rezza G, Cassone A. 2007. An outbreak of chikungunya fever in the province of Ravenna, Italy. *Euro Surveill* 12:E070906 070901.
 42. Grandadam M, Caro V, Plumet S, Thiberge JM, Souares Y, Faillox AB, Toulon HJ, Budelot M, Cossart D, Leparac-Goffart I, Despres P. 2011. Chikungunya virus, southeastern France. *Emerg Infect Dis* 17:910–913. <http://dx.doi.org/10.3201/eid1705.101873>.
 43. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IR, Teng HJ, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GW, Golding N, Hay SI. 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife* 4:e08347. <http://dx.doi.org/10.7554/eLife.08347>.
 44. D'Ortenzio E, Grandadam M, Ballezardier E, Jaffar-Bandjee MC, Michault A, Brottet E, Bavielle M, Filleul L. 2011. A226V strains of Chikungunya virus, Reunion Island, 2010. *Emerg Infect Dis* 17:309–311. <http://dx.doi.org/10.3201/eid1702.101056>.
 45. Diallo M, Thonnon J, Traore-Lamizana M, Fontenille D. 1999. Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *Am J Trop Med Hyg* 60:281–286.
 46. Peyrefitte CN, Rousset D, Pastorino BA, Pouillot R, Bessaud M, Tock F, Mansaray H, Merle OL, Pascual AM, Paupy C, Vessiere A, Imbert P, Tchendjou P, Durand JP, Tolou HJ, Grandadam M. 2007. Chikungunya virus, Cameroon, 2006. *Emerg Infect Dis* 13:768–771. <http://dx.doi.org/10.3201/eid1305.061500>.
 47. Paupy C, Kassa F, Caron M, Nkoghe D, Leroy EM. 2012. A chikungunya outbreak associated with the vector *Aedes albopictus* in remote villages of Gabon. *Vector Borne Zoonotic Dis* 12:167–169. <http://dx.doi.org/10.1089/vbz.2011.0736>.
 48. Moyen N, Thiberville SD, Pastorino B, Nougaiere A, Thirion L, Mombouli JV, Dimi Y, Leparac-Goffart I, Capobianchi MR, Lepfoundzou AD, de Lamballerie X. 2014. First reported chikungunya fever outbreak in the republic of Congo, 2011. *PLoS One* 9:e115938. <http://dx.doi.org/10.1371/journal.pone.0115938>.
 49. Chen R, Wang E, Tssetsarkin KA, Weaver SC. 2013. Chikungunya virus 3' untranslated region: adaptation to mosquitoes and a population bottleneck as major evolutionary forces. *PLoS Pathog* 9:e1003591. <http://dx.doi.org/10.1371/journal.ppat.1003591>.
 50. Diaz-Gonzalez EE, Kautz TF, Dorantes-Delgado A, Malo-Garcia IR, Laguna-Aguilar M, Langsjoen RM, Chen R, Auguste DI, Sanchez-Casas RM, Danis-Lozano R, Weaver SC, Fernandez-Salas I. 2015. First report of *Aedes aegypti* transmission of chikungunya virus in the Americas. *Am J Trop Med Hyg* 93:1325–1329. <http://dx.doi.org/10.4269/ajtmh.15-0450>.
 51. Kaariainen L, Ahola T. 2002. Functions of alphavirus nonstructural proteins in RNA replication. *Prog Nucleic Acid Res Mol Biol* 71:187–222. [http://dx.doi.org/10.1016/S0079-6603\(02\)71044-1](http://dx.doi.org/10.1016/S0079-6603(02)71044-1).
 52. Saxton-Shaw KD, Ledermann JP, Borland EM, Stovall JL, Mossel EC, Singh AJ, Wilusz J, Powers AM. 2013. O'nyong nyong virus molecular determinants of unique vector specificity reside in non-structural protein 3. *PLoS Negl Trop Dis* 7:e1931. <http://dx.doi.org/10.1371/journal.pntd.0001931>.
 53. Lastarza MW, Grakoui A, Rice CM. 1994. Deletion and duplication mutations in the C-terminal nonconserved region of Sindbis virus nsP3: effects on phosphorylation and on virus replication in vertebrate and invertebrate cells. *Virology* 202:224–232. <http://dx.doi.org/10.1006/viro.1994.1338>.
 54. Galbraith SE, Sheahan BJ, Atkins GJ. 2006. Deletions in the hypervariable domain of the nsP3 gene attenuate Semliki Forest virus virulence. *J Gen Virol* 87:937–947. <http://dx.doi.org/10.1099/vir.0.81406-0>.
 55. Faria NR, Azevedo RDS, Kraemer MU, Souza R, Cunha MS, Hill SC,

- These J, Bonsall MB, Bowden TA, Rissanen I, Rocco IM, Nogueira JS, Maeda AY, Vasami FG, Macedo FL, Suzuki A, Rodrigues SG, Cruz AC, Nunes BT, Medeiros DB, Rodrigues DS, Nunes Queiroz AL, da Silva EV, Henriques DF, Travassos da Rosa ES, de Oliveira CS, Martins LC, Vasconcelos HB, Casseb LM, Smith DDB, Messina JP, Abade L, Lourenco J, Carlos Alcantara L, de Lima MM, Giovanetti M, Hay SI, de Oliveira RS, Lemos PDS, de Oliveira LF, de Lima CP, da Silva SP, de Vasconcelos JM, Franco L, Cardoso JF, Vianez-Junior JL, Mir D, Bello G, Delatorre E, Khan K, Creatore M, Coelho GE, de Oliveira WK, Tesh R, Pybus OG, Nunes MR, Vasconcelos PF. 2016. Zika virus in the Americas: early epidemiological and genetic findings. *Science* 352:345–349. <http://dx.doi.org/10.1126/science.aaf5036>.
56. Naresh Kumar CV, Sivaprasad Y, Sai Gopal DV. 2016. Genetic diversity of 2006–2009 Chikungunya virus outbreaks in Andhra Pradesh, India, reveals complete absence of E1:A226V mutation. *Acta Virol* 60:114–117. http://dx.doi.org/10.4149/av_2016_01_114.
57. Parashar D, Amdekar S, More A, Patil P, More R, Babu VR. 2015. Chikungunya fever outbreak in Guntur, Andhra Pradesh, India. *Indian J Med Res* 142(Suppl):S111–S115. <http://dx.doi.org/10.4103/0971-5916.176639>.
58. Sumathy K, Ella KM. 2012. Genetic diversity of Chikungunya virus, India 2006–2010: evolutionary dynamics and serotype analyses. *J Med Virol* 84:462–470. <http://dx.doi.org/10.1002/jmv.23187>.
59. Nitatattana N, Kanjanopas K, Yoksan S, Satimai W, Vongba N, Langdatsuan S, Nakgoi K, Ratchakum S, Wauquier N, Souris M, Auewarakul P, Gonzalez JP. 2014. Long-term persistence of Chikungunya virus neutralizing antibodies in human populations of North Eastern Thailand. *Virol J* 11:183. <http://dx.doi.org/10.1186/1743-422X-11-183>.
60. Salje H, Cauchemez S, Alera MT, Rodriguez-Barraquer I, Thaisomboonsuk B, Srikiatkachorn A, Lago CB, Villa D, Klungthong C, Tac-An IA, Fernandez S, Velasco JM, Roque VG, Jr, Nisalak A, Macareo LR, Levy JW, Cummings D, Yoon IK. 2016. Reconstruction of 60 years of chikungunya epidemiology in the Philippines demonstrates episodic and focal transmission. *J Infect Dis* 213:604–610. <http://dx.doi.org/10.1093/infdis/jiv470>.
61. Yoon IK, Alera MT, Lago CB, Tac-An IA, Villa D, Fernandez S, Thaisomboonsuk B, Klungthong C, Levy JW, Velasco JM, Roque VG, Jr, Salje H, Macareo LR, Hermann LL, Nisalak A, Srikiatkachorn A. 2015. High rate of subclinical chikungunya virus infection and association of neutralizing antibody with protection in a prospective cohort in the Philippines. *PLoS Negl Trop Dis* 9:e0003764. <http://dx.doi.org/10.1371/journal.pntd.0003764>.
62. Galatas B, Ly S, Duong V, Baisley K, Nguon K, Chan S, Huy R, Ly S, Sorn S, Som L, Buchy P, Tarantola A. 2016. Long-lasting immune protection and other epidemiological findings after chikungunya emergence in a Cambodian rural community, April 2012. *PLoS Negl Trop Dis* 10:e0004281. <http://dx.doi.org/10.1371/journal.pntd.0004281>.
63. AbuBakar S, Sam IC, Wong PF, MatRahim N, Hooi PS, Roslan N. 2007. Reemergence of endemic Chikungunya, Malaysia. *Emerg Infect Dis* 13:147–149. <http://dx.doi.org/10.3201/eid1301.060617>.
64. Noridah O, Paranthaman V, Nayar SK, Masliza M, Ranjit K, Norizah I, Chem YK, Mustafa B, Kumarasamy V, Chua KB. 2007. Outbreak of chikungunya due to virus of Central/East African genotype in Malaysia. *Med J Malaysia* 62:323–328.
65. Chua KB. 2010. Epidemiology of chikungunya in Malaysia: 2006–2009. *Med J Malaysia* 65:277–282.
66. Sam IC, Chan YF, Chan SY, Loong SK, Chin HK, Hooi PS, Ganeswrie R, Abubakar S. 2009. Chikungunya virus of Asian and Central/East African genotypes in Malaysia. *J Clin Virol* 46:180–183. <http://dx.doi.org/10.1016/j.jcv.2009.07.016>.
67. Maha MS, Susilarini NK, Hariastuti NI, Subangkit. 2015. Chikungunya virus mutation, Indonesia, 2011. *Emerg Infect Dis* 21:379–381. <http://dx.doi.org/10.3201/eid2102.141121>.
68. Sy AK, Saito-Obata M, Medado IA, Tohma K, Dapat C, Segubre-Mercado E, Tandoc A III, Lupisan S, Oshitani H. 2016. Molecular characterization of chikungunya virus, Philippines, 2011–2013. *Emerg Infect Dis* 22:887–890. <http://dx.doi.org/10.3201/eid2205.151268>.
69. Contevelle LC, Zanella L, Marin MA, Filippis AM, Nogueira RM, Vicente AC, Mendonca MC. 2016. Phylogenetic analyses of chikungunya virus among travelers in Rio de Janeiro, Brazil, 2014–2015. *Mem Inst Oswaldo Cruz* 111:347–348. <http://dx.doi.org/10.1590/0074-02760160004>.
70. Tssetsarkin KA, McGee CE, Volk SM, Vanlandingham DL, Weaver SC, Higgs S. 2009. Epistatic roles of E2 glycoprotein mutations in adaptation of chikungunya virus to *Aedes albopictus* and *Ae. aegypti* mosquitoes. *PLoS One* 4:e6835. <http://dx.doi.org/10.1371/journal.pone.0006835>.
71. Ganushkina LA, Patraman IV, Rezza G, Migliorini L, Litvinov SK, Sergiev VP. 2016. Detection of *Aedes aegypti*, *Aedes albopictus*, and *Aedes koreicus* in the area of Sochi, Russia. *Vector Borne Zoonotic Dis* 16:58–60. <http://dx.doi.org/10.1089/vbz.2014.1761>.
72. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.