Infection, Genetics and Evolution 9 (2009) 16–23

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Evolutionary rates and timescale comparison of Chikungunya viruses inferred from the whole genome/E1 gene with special reference to the 2005–07 outbreak in the Indian subcontinent

Sarah S. Cherian *, Atul M. Walimbe, Santosh M. Jadhav, Swati S. Gandhe, Supriya L. Hundekar, Akhilesh C. Mishra, Vidya A. Arankalle

National Institute of Virology, 20-A, Dr. Ambedkar Road, Post Box No. 11, Pune 411001, Maharashtra State, India

ARTICLE INFO

Article history:
Received 30 April 2008
Received in revised form 17 September 2008
Accepted 22 September 2008
Available online 1 October 2008

Keywords:
Chikungunya virus
Evolutionary rates
Divergence times
Bayesian
MCMC
tMRCA
India

ABSTRACT

Chikungunya (CHIK) virus reemerged during 2005–07 as an important pathogen causing massive disease outbreaks affecting India and several countries of the Indian Ocean. Knowledge of the evolutionary rates and divergence times of the CHIK virus may help to better understand the disease epidemiology. Considering the limited availability of such information, we estimated the substitution rates and the ancestral times for all the CHIK genotypes and also the time to the most recent common ancestor (tMRCA) of the 2005–07 isolates. Using whole genomes and partial E1 gene datasets, we applied the Bayesian Markov Chain Monte Carlo (MCMC) framework that explicitly accounts for lineage-specific evolutionary rates through the use of 'relaxed' molecular clock models. Under a constant population relaxed clock model, the evolutionary timescale of CHIK viruses in this study was estimated to be in the last 300 years. The progenitor of the 2005–07 viruses was found to have existed around 9 years ago, and to have originated from Central Africa. The presence of a strain in India in 2000 that bears 99% identity with a Ugandan strain of 1982, which correlates with the tMRCA of the Indian and Indian Ocean isolates, confirms our earlier report that the progenitor of the 2005–07 isolates originates from Uganda's neighbourhood. The 'A226V' mutation that existed in the Indian Ocean isolates since late 2005 was found to occur only in the 2007 isolate from India. The study confirms the epidemiological data, specifically with regard to the re-emergence of CHIKV and throws light on the evolutionary dynamics of CHIK viruses.

1. Introduction

The Chikungunya virus epidemic, which resurged after a gap of about 32 years in India (Arankalle et al., 2007) in January 2006, continued intense activity in several states of India up to August 2007 (http://www.who.int/ith/maps/en/dtd21/04/08). The virus was epidemic on the Reunion islands in the Indian Ocean and other neighbouring islands of Seychelles, Madagascar, Maurice, Mauritius, Mayotte (Chastel, 2005; Beltrame et al., 2007) during the period from February 2005 to June 2006. The *Aedes albopictus* mosquito was the presumed vector in Reunion while in India it was predominantly *Aedes aegypti*. (Yergolkar et al., 2006; Pialoux et al., 2007). The clinical manifestations are marked by severe joint pains, fever, headache, rash, nausea, vomiting, polyarthralgia, etc. Severe and unusual forms not previously described in the literature have also appeared. There is limited information regarding the evolutionary rates and divergence times of the CHIK virus, which may help to better understand the disease epidemiology.

CHIK, a positive stranded RNA virus, is a member of the *Alphavirus* genus of the family *Togaviridae*. It belongs to the Semliki Forest Virus (SFV) group, closely related to the O'nyong-nyong (ONN) virus. According to the genomic organization of other Alphaviruses, the genome of CHIK is considered to be: 5′ nsP1-nsP2-nsP3-nsP4-junction region-C-E3-E2-6K-E1-poly (A) 3′. The coding sequence (CDS) comprising of the nonstructural and structural polyproteins is of approximately 11 kb (Strauss and Strauss, 1986).

The virus was first isolated during an outbreak in Taznania in 1952 (Ross, 1956). Between 1960s and 1980s, the virus was isolated repeatedly from various countries in central and southern Africa as well as from Senegal and Nigeria in western Africa (Kokernot et al., 1965; Macrae et al., 1971). CHIK appears to have spread to other parts of the world from Africa to cause pandemics in Asian tropics (Powers et al., 2000). CHIK activity in Asia has been documented since its isolation in Bangkok, Thailand in 1958 (Hammon et al., 1960). Other Asian and southeast Asian countries affected are Myanmar, Philippines, Malaysia, Pakistan and the

* Corresponding author. Tel.: +91 20 26127301; fax: +91 20 26122669.
E-mail address: sarahcherian@icmr.org.in (S.S. Cherian).

1567-1348/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2008.09.004

Pacific islands (Jupp and McIntosh, 1988). The aforementioned report also documented that India had a history of CHIK epidemics from 1824 to 1923. It was also suggested that CHIK became established in southeast Asia during the late 1950s and early 1960s. The first outbreak in India was recorded in Kolkata in 1963 (Shah et al., 1964), followed by epidemics during 1964 in the east coast areas like Chennai, Pondicherry, Vellore, Vishakapattanam (Rao, 1966). Investigations in Nagpur in 1965 showed that the incidence was as high as 40–70% in certain wards and mortality was negligible (Rodrigues et al., 1972). Sporadic cases were recorded in 1973 in Barsi, Maharashtra (Padbidri and Gnaneswar, 1979).

Phylogenetic studies based on the partial E1 gene have shown the delineation of two distinct CHIK virus lineages, one containing all isolates from western Africa and the second comprising all south, east and central African strains, as well as isolates from Asia (Powers et al., 2000; Pastorino et al., 2004). It is also reported that the strains from Africa and Asia differ biologically and a distinct Asian subgroup is also evident (Jupp and McIntosh, 1988). The strains causing the epidemics in India during 1963 and 1973 clustered into the Asian subgroup (Arankalle et al., 2007).

Interestingly, a mosquito isolate (MH4_2000) from Yawat town in Maharashtra (Mourya et al., 2001) that was recently sequenced was identified as the African genotype (Yergolkar et al., 2006). In addition, the full genome of five CHIK isolates from the 2006 epidemic, representative of different Indian states, were sequenced and compared with the Reunion isolates. Sequence analysis was carried out to understand the association if any of the unique mutations with the explosive nature of the outbreak (Arankalle et al., 2007). The relatedness of the Indian isolates with the Reunion island isolates was determined and the virus was found to be the so-called East-Central-South-Africa (ECSA) genotype. Further, no evidence of positive selection pressure or the recombination of the Asian and African genotypes was reported. With sporadic cases of CHIK still being reported from certain parts of the country, it is essential to understand the evolutionary mechanism of this virus including the substitution rates, divergence times and the time for the most recent common ancestor (tMRCA) of the 2005–07 isolates.

The only report (Powers et al., 2000) based on a phylogenetic analysis of partial E1 gene sequences estimates the rate of CHIK evolution based on six sister pairs of isolates as 6×10^{-4} substitutions per nucleotide per year (standard deviation 4×10^{-4}). Further, on the basis of phylogenetic methods, the Asian genotype was estimated to have emerged between 50 and 310 years ago while the West and East African genotypes diverged between 100 and 840 years ago. Due to limited number of sequences available and analyzed at that point of time, one expects less reliability of the estimates and wider confidence intervals. Hence it becomes ever more vital to reanalyze now available larger data sets using more accurate dating methods based on rigorous hypothesis testing and statistical approaches.

Analytical methods for estimating evolutionary rates and divergence times have been recently developed that explicitly account for lineage-specific rates through the use of "relaxed" molecular clock models (Drummond et al., 2006). Using these methods, we report here the results based on analysis of whole genomes as well as partial E1 protein sequences from the isolates belonging to the different genotypes. Investigating the evolutionary patterns of the virus would help better understand the epidemiology of the disease.

2. Materials and methods

Partial E1 gene sequences ($n = 65$) of length 1044 nt and 27 whole genomes (CDS region only, 11166 nt) of CHIK were analyzed in the present study. These include thirteen whole genome sequences of isolates of 2006–07 from several states of India sequenced at the National Institute of Virology. All the virus sequences had a known sampling time. The description of the sequences used herein is available as Supplementary Table 1. Multiple sequence alignments were carried out using ClustalX 1.83 (Thompson et al., 1997). MEGA ver. 3.1 (Kumar et al., 2004) was used for calculation of the percent identities.

To select the best fitting nucleotide substitution model, different models of evolution were fitted to the data by using PAUP*, ver. 4.0b10 (Swofford, 2003). The model selected by Akaike Information Criterion (AIC) as implemented in Model Test 3.7 (Posada and Crandall, 1998) was treated as the best model. For the whole genome set this was found to be GTR (General Time Reversible) + I (proportion of Invariant sites) while for the E1 gene, it was GTR + $\Gamma 4$ (gamma distributed rate variation with four rate categories). Maximum likelihood (ML) trees were inferred using Treefinder 2007 (Jobb et al., 2004) and reliability of the trees were tested by bootstrapping with 500 replicates.

2.1. Molecular clock testing

PAML (Yang, 2007) was used to study three models of branch rate variation viz. Different Rate (DR), Single Rate (SR), and the Single Rate Dated Tip (SRDT) for the two data sets. The mid point rooted ML tree was used to fit the SR and SRDT models while the unrooted ML tree was used to fit the DR model. The SR model assumes that all branches have the same rate of evolution over time, i.e. a strict molecular clock, while in the DR model each branch has its own rate. The SRDT model is an extension of the SR model and relaxes the assumption of contemporaneous tips (Rambaut, 2000). The molecular clock hypothesis (SR vs. DR) and (SRDT vs. DR) was tested using the likelihood ratio test (LRT).

2.2. Estimation of evolutionary rates and the tMRCA

The Bayesian Markov Chain Monte Carlo (MCMC) approach available in BEAST 1.4.6 package (Drummond and Rambaut, 2007) was employed to estimate the rates of evolutionary change, the divergence times and the tMRCA of the CHIK viruses. The strict molecular clock and the relaxed clock (uncorrelated exponential, uncorrelated lognormal) (Drummond et al., 2006) models were implemented for both the datasets, whole genome and the partial E1 gene. For each case the constant size, exponential, logistic and expansion growth demographic population models were applied. The Bayesian skyline plot (Drummond et al., 2005) was used to infer the possible demographic population model that may best fit the data. Uncertainty in the parameter estimates is reflected in the 95% highest probability density (HPD) limits.

At least three independent MCMC chains were run for 10,000,000 generations, sampling every 1000th generation. The burnin was set to 10% generations during combination of the runs in LogCombiner available in BEAST. Tracer 1.4 (Rambaut and Drummond, 2007) was used to evaluate and ensure the convergence of the mixing in terms of the effective sample size (ESS) values, density plots and trace plots. Model comparison was done by calculation of Bayes factor on the basis of the relative marginal likelihoods of the models under comparison (Suchard et al., 2001). The maximum clade credibility tree was generated using Tree Annotator available in BEAST and FigTree (http://tree.bio.ed.ac.uk/) was used for visualization of the annotated trees.

Table 1
Testing of the molecular clock hypothesis models for whole genomes and partial E1 gene sequences.

Data set	Model	χ^2	df	p value	Result
Whole genome (n = 27, 11166 nt)	SR(H_0) vs. DR(H_1)	118.65	25	<0.001	Accept DR
	SRDT(H_0) vs. DR(H_1)	97.48	24	<0.001	Accept DR
Partial E1 gene (n = 65, 1044 nt)	SR(H_0) vs. DR(H_1)	191.45	63	<0.001	Accept DR
	SRDT(H_0) vs. DR(H_1)	163.88	62	<0.001	Accept DR

3. Results

3.1. Tests of molecular clock hypothesis

To test the molecular clock hypothesis, the LRT was applied to compare the fit of the SR and SRDT models against the DR model. The SR model does not accommodate for the different dates of sampling of the isolates. The LRT comparing the SR model against the DR model thus indicates the evidence for rate constancy among lineages in the CHIK phylogeny. On the other hand, the SRDT makes accommodation for the temporal sampling of the isolates. If the SRDT model is not significantly worse than the DR model, it indicates that this model adequately describes the substitution process.

Table 1 gives the results of the LRTs for the whole genome and E1 gene datasets. It can be seen that in both the cases, the SR as well as the SRDT models are strongly rejected ($P < 0.001$) in comparison to the DR model.

3.2. Estimation of evolutionary rates and the tMRCA

Table 2 represents estimates of different parameters obtained under different combinations of branch rate variation models and population growth models that were observed to have converged for the whole genome as well as the E1 gene data sets. It can be seen that the marginal log likelihoods as well as the posterior probabilities for the relaxed clock (uncorrelated exponential) model are much higher than those for the strict clock model for both the datasets. The Bayesian skyline plots (Fig. 1a and b) indicate a near constant phase over a period of approximately 50 years (1937–1987), followed by a decline phase for roughly 15–20 years and a rapid growth phase over the period between 2004 and 2007. Among the two population demographies, the exponential growth model was better in terms of higher posterior probability. However, the marginal likelihoods were not significantly different in these two models as was evident from the Bayes factor. Further, on the basis of the skyline plots, that showed a constant population over a relatively longer time frame, the constant population size, relaxed clock model (Table 2), was selected to be the best fit model for the currently available data.

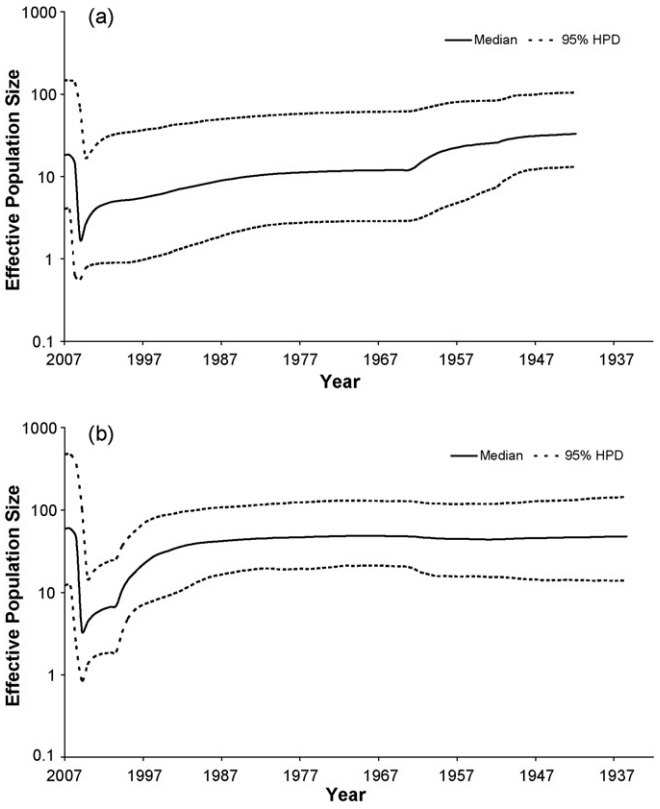

Fig. 1. Bayesian skyline plot under the relaxed clock constant population model for (a) whole genomes and (b) partial E1 gene sequences.

The Bayesian MCMC maximum a posteriori (MAP) tree depicting the phylogeny under the best-fit relaxed clock (uncorrelated exponential, constant population growth) model for the whole genomes and the E1 gene data set, is shown in Fig. 2a and b respectively. The tree topology for the whole genome data set (Fig. 2a) is identical to that reported earlier (Arankalle et al., 2007) with all the monophyletic nodes showing 100% posterior support. In case of the E1 tree, the topology is identical with the exception that the Asian and East African sub-lineages swapped their positions with a low posterior support of 66%. This discrepancy is not observed in the MCMC tree for the E1 gene data set under the strict clock model. Thus the partial loss of the branching topology in the E1 gene sequences may be attributed to the more parameter-rich relaxed clock model being fitted to shorter length sequences.

Table 2
Estimates of substitution rates and root age for whole genomes and partial E1 gene sequences using BEAST.

Dataset	Model	Marginal likelihood	Posterior probability	Root age (years)	Mean clock rate (sub/site/year) × 10^{-4}
Whole genome (n = 27, 11166 nt)	Strict constant population	−26296.36 ± 0.12	−26410.0 (−26430.0, −26400.0)	893.0 (594–1285)	1.7 (1.0–2.3)
	Strict exponential growth	−26297.03 ± 0.13	−26400.0 (−26410.0, −26390.0)	774.0 (543.6–1045.5)	1.9 (1.3–2.5)
	Relaxed constant population	−26258.32 × 0.13	−26340.0 (−26360.0, −26320.0)	167.3 (71.4–299.3)	8.8 (3.3–14.8)
	Relaxed exponential growth	−26259.11 ± 0.15	−26310.0 (−26330.0, −26280.0)	129.4 (60.5–225.8)	12.1 (5.0–20.3)
E1 gene (n = 65, 1044 nt)	Strict constant population	−3750.45 ± 0.18	−4036.8 (−4059.5, −4015.7)	301.7 (227.1–385.2)	4.4 (3.3–5.5)
	Strict exponential growth	−3750.07 ± 0.17	−4026.9 (−4050.6, −4005.2)	299.6 (226.8–382.0)	4.5 (3.5–5.7)
	Relaxed constant population	−3710.91 ± 0.20	−3958.2 (−3990.2, −3928.2)	140.1 (79.4–215.3)	8.3 (5.3–11.7)
	Relaxed exponential growth	−3710.11 ± 0.18	−3939.9 (−3974.8, −3906.4)	155.6 (79.1–262.8)	8.4 (5.0–12.3)

The mean substitution rate of the 27 complete genomes under the best-fit model is 8.8×10^{-4} subs/site/year with 95% HPD limits of $3.3–14.8 \times 10^{-4}$. On the other hand, the mean evolutionary rate observed in case the E1 gene data set is 8.3×10^{-4} subs/site/year with a narrower range of the 95% HPDs viz. $5.3–11.7 \times 10^{-4}$ (Table 2).

The estimates for the time to the MRCA (tMRCA) of the different genotypes of CHIK, depicted as the mean age of the root, under the best fit model for the whole genome dataset is 167 years with the upper bound of 299 years while for the E1 gene data set, 140 years with a maximum limit of 215 years. Under the strict molecular clock, estimates for the tMRCA fall in the range 544 years to 1285 years for the 27 whole genomes while these are between 227 and 385 years for the E1 gene data set.

The mean age of the node indicating the divergence of the East/South African and the Asian sub-lineages in the best fit model with respect to the whole genome, is 88 years with HPD limits 61–124 years, while for the E1 gene dataset it is seen to be 73 years with HPD limits 56–95 years (Fig. 2a and b respectively). It is noteworthy that these values are comparable for both the data sets.

The common ancestor of the MH4_2000 strain and the isolates corresponding to the 2005–07 outbreak in India/Indian Ocean had a mean age of 27 years (95% HPD: 11–49 years) based on the whole genome dataset. The corresponding estimate for the E1 gene sequences relating to the divergence of the Central African (including isolates from Democratic Republic of Congo, Uganda, Cameroon and the lone isolate from India viz MH4_2000) and the 2005–07 isolates is 35 years (95% HPD: 28–45 years). It is noteworthy that the estimate for the E1 gene sequences has been made from a larger sample size as compared to that for whole genomes, resulting in narrower HPD limits. Thus, it can be stated that the node defining the divergence of the Central African strains may have evolved around 28–45 years ago.

The tMRCA for the 2005–07 isolates is about 9 years, interestingly from either of the datasets studied, with comparable estimates of 95% HPD (3–16 years). The Cameroon E1 gene sequence in 2006 (Peyrefitte, 2007) clustered with the isolates from the Democratic Republic of Congo in 2000.

Interestingly, the isolates of the 2005–07 epidemic, geographically cluster into two distinct groups with 100% posterior support. The isolates of US travelers from India (Lanciotti et al., 2007) during the period of the epidemic cluster with the other isolates from India while those from travelers to the Reunion islands accordingly cluster with the Indian Ocean isolates. Further the eight new isolates from India, including the 2007 isolate, also cluster with the earlier 2006 Indian isolates.

In order to understand the diversity of the Indian and the Indian Ocean isolates the percent nucleotide and amino acid divergence within and between themselves and with respect to the other closely related isolates were determined. The divergence over the whole genomes within the Indian isolates in the 2005–07 period is 0.11% while it is 0.04% for the Indian Ocean isolates (data not shown). Thus, apparently there is an increase in the divergence within the Indian isolates over the period 2005–07. We have reported earlier (Arankalle et al., 2007) that irrespective of the geographical location, the 2005–06 isolates were very closely related having 99.9% nucleotide identity. This identity is retained (99.8%) even with additional isolates in both the groups.

Among all the genes (Supplementary Table 2), maximum divergence at nucleotide level (0.32%) between the two groups is observed in the E1 region. Furthermore, it is seen that except for

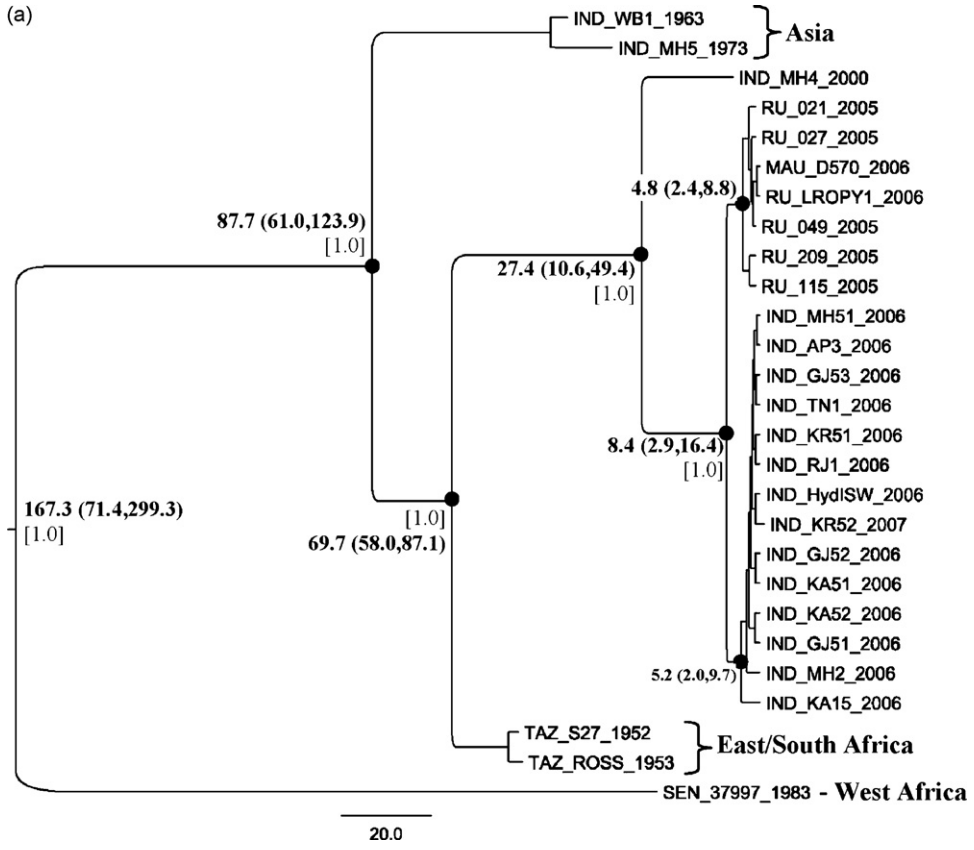

Fig. 2. Bayesian MCMC tree for (a) whole genomes and (b) partial E1 gene sequences under the relaxed clock constant population model. Also shown are the posterior probabilities (in square brackets), estimated ages (bold letters).

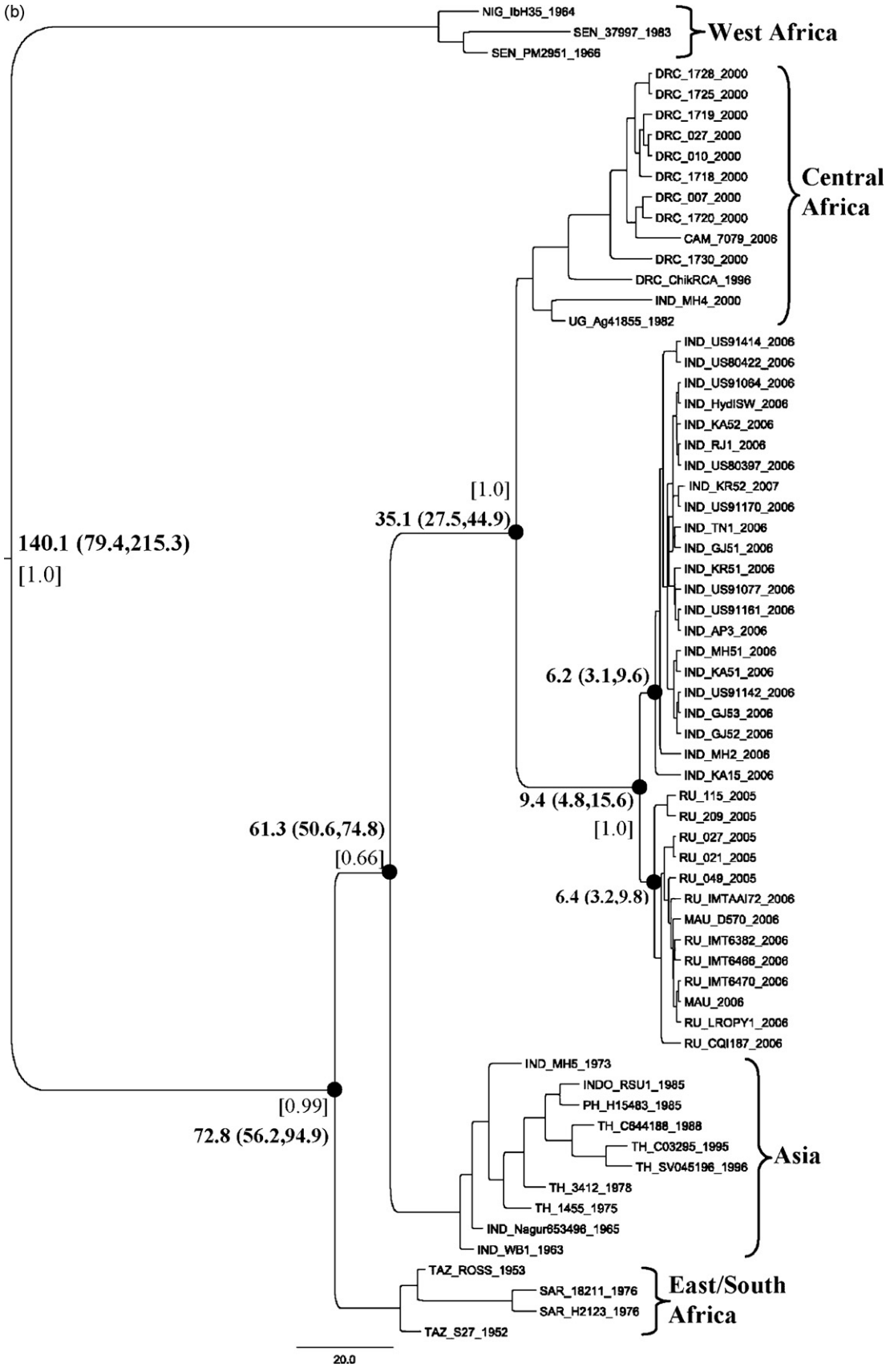

Fig. 2. (*Continued*).

the nsP4 and the E1 gene, in all cases the Indian Ocean isolates have a higher percent identity with the MH4_2000 and the East/South African strains when compared to the Indian sequences. There are eight nucleotide sites viz. 5248C/T, 9633T/C, 358T/C, 459A/C → 128K/T (Nsp1), 1203T/C → 376 M/T (Nsp1), 7633T/C → 2497S/P (Capsid), 8910T/C and 11127T/C (numbering w.r.t S27 strain; nucleotide or amino acid in Indian/Indian Ocean isolate and → indicates an amino acid change) distinguishing the Indian isolates from the Indian Ocean ones. Of these, the first two sites in the Indian isolates are shared with the MH4_2000 strain while the other six are unique. Alternatively, the Indian Ocean isolates have the respective two sites as unique and the other six sites shared with MH4_2000 strain.

Within the Indian sublineage, the earliest, January 2006, isolate from Karnataka (IND_KA15), maintains three nucleotides (3724A, 10314C and 10743G) as in the East/South African strains and Reunion isolates. Of these, 10314C is also observed in two more Indian isolates, IND_MH51 and IND_KA51 isolated in August 2006, from Maharashtra and Karnataka respectively. Two substitutions, 9565C and 10626C → 211N (E1) were interestingly noted only in the IND_KA15 and IND_GJ51 (Gujarat, September 2006) isolates that were otherwise maintained as T and A respectively in all other Indian as well as Reunion isolates. As a result, the IND_KA15 isolate was outgrouped from the other Indian isolates (Fig. 2). 10743G as in IND_KA15 is also maintained in a February 2006 isolate from Maharashtra, IND_MH2. Another isolate of February 2006, from Andhra Pradesh, IND_AP3, maintained one non synonymous substitution, 7645G → 27 V (Capsid), as in the East/South African strain and Reunion isolates and so also in a isolate from Nagpur, Maharashtra (IND_MH51) of August 2006. Hence, the two isolates grouped together (Fig. 2a) with high posterior support (85%). A single substitution, 5577C → 501S (Nsp3), was shared between isolates IND_KA51, Karnataka, August 2006 and IND_GJ52, Gujarat, September 2006. Three substitutions, 268T, 343C and 8965T → 142Y (E2) were shared between IND_GJ51, Gujarat, September 2006 and IND_KA52, Karnataka, October 2006. The two pairs also clustered together (Fig. 2a) with posterior support of at least 65%. The other pairs had poor posterior supports. The IND_KR52, Kerala, 2007 isolate, showed eight amino acid differences with the IND_KR51, Kerala, 2006 isolate of which four viz. 48A and 539S (Nsp2); 252Q (E2); and V226 (E1) were unique to the IND_KR52, while four (Nsp2: C120; E2: L307, P313; E1:Q107) were unique to the IND_KR51 isolate.

In case of the partial E1 gene, compared to the Indian Ocean isolates, the Indian isolates had a higher amino acid identity with both MH4_2000 (99.3% vs. 98.9%) and the East/South African strains (98.9% vs. 98.6%). The mutation A226V that was observed in the IND_KR52 and has also been recently reported (Santosh et al., 2008; Kumar et al., 2008), was observed in the Reunion isolates after September 2005 in the latter phase of their epidemic (Schuffenecker et al., 2006).

4. Discussion

In the present study, we analyzed two different data sets. Considering the recent availability of a number of whole genomes ($n = 27$), it was thought apt to estimate evolutionary time scales using this data set. Further, to investigate whether the E1 gene would suffice in terms of phylogenetic signal and estimates of dates and rates, we applied various clock models on the enhanced E1 data set ($n = 65$) also.

The strict molecular clock was rejected by the likelihood analysis in favor of a relaxed clock model indicative of an evolutionary rate variation in different lineages of CHIK. The MCMC analysis showed that the best fit model was the relaxed, uncorrelated exponential model with constant population size. The substitution rates of 8.8×10^{-4} subs/site/year observed under the best fit model are comparable with the estimate for the CHIK virus (Powers et al., 2000) and higher when compared to the known rates for other alphaviruses (Weaver et al., 1992, 1997). Further, the estimates of the mean age of the root of CHIK genomes under the strict molecular model were close to the earlier estimates of 100–840 years (Powers et al., 2000), indicating that the earlier approaches correlate to a strict molecular clock. Interestingly, our Bayesian analysis revealed that the CHIK strains may have a relatively recent ancestry, of about 300 years.

The estimates of the tMRCA for different sub-lineages of the CHIK viruses were also investigated in this study. The estimate that the East/South African and the Asian sub- lineages could have diverged around 125 years ago corresponds to CHIK activity in Asia in the early 1880 s correlating well with epidemiological evidence in India and Asia. The tMRCA of the 2005–07 isolates being in the range from 3 to 16 years notably corresponds to the period of activity of the Central African strains (Pastorino et al., 2004; Powers et al., 2000) and it may follow that the present epidemic has arisen from the population of the strains circulating therein. The existence of the Cameroon E1 gene sequence in 2006 clustering with the isolates from the Democratic Republic of Congo in 2000 is indicative of the relative stability of the CHIK viruses circulating in that region which is distinct from the strain circulating in the Indian Ocean and India. A recent report (Peyrefitte et al., 2008) has shown that the Gabon sequence of 2007 also clusters with the Cameroon-2006 sequence.

We had shown in our earlier paper (Arankalle et al., 2007) that both Indian and the Indian Ocean isolates bear highest nucleotide identity with the MH4_2000 strain (98.1% and 98.2%), followed by Uganda-82 (97.7% and 97.8% respectively) and East/South African strains (97.1% for both). Lower percent identities are obtained with the isolates from DRC (96.7% and 96.8%) though these are samples of the year 2000. The fact that the 2005–07 epidemic isolates are more closely related to the MH4_2000 and the Uganda82 strains implies that there is reason to ascertain that the present isolates have evolved from a Uganda-like strain. Further, eight nucleotide sites were found to be responsible for delineating the Indian Ocean and Indian isolates and on the whole, the Indian isolates were found to have accumulated more changes with respect to the MH4_2000 strain. Within the Indian sublineage, subclusters were indicative of the role of temporal factors more than geographical ones.

The A226V mutation in the E1 gene deserves special mention. In the Reunion, this mutation was accompanied by a concomitant spectacular increase in the frequency of CHIK infections and was thought to be responsible for the increased transmissibility of the virus and higher epidemic potential (Schuffenecker et al., 2006). Laboratory experiments provided evidence for the importance of this mutation as a cause of significant increase in the infectivity of CHIK virus for *Ae. albopictus* and not *Ae. aegypti* mosquitoes (Vazeille et al., 2007; Tsetsarkin et al., 2007). It is important to note here that *albopictus* and *aegypti* have been shown to be the main vectors in the Indian Ocean islands (Schuffenecker et al., 2006) and India (Yergolkar et al., 2006) respectively. Large epidemics in several Indian states (NVBDCP, 2007) in the absence of A226V mutation in the 2006 isolates (Arankalle et al., 2007) corroborate the involvement of *Ae. aegypti*. The detection of A226V mutation in the June 2007 isolate from Kerala is an important observation, considering the fact that *Ae. albopictus* is the predominant vector in this state (Kumar et al., 2008). However, the maximum suspected CHIK cases were observed before this time point (NVBDCP, 2007) and hence indicative of the acquisition of the A226V mutation as a

probable function of time. Another recent report (Lamballerie et al., 2008) also describes the acquisition of the single adaptive mutation, A226V as evolutionary convergence.

The possibility of two independent introductions to India and islands of the Indian Ocean or the first introduction into either of the regions cannot be ruled out. The presence of the MH4_2000 strain in India (bearing 99% identity with the Uganda-1982 strain) that correlates with the tMRCA of the 2005–07 isolates, suggests the first introduction of a Uganda-like strain in India around the year 2000 without an epidemic potential. The earliest Indian isolates did maintain some of the substitutions as in the Reunion isolates bringing out a possibility of a first introduction into the islands of the Indian Ocean. However, considering the A226V mutation as a function of time and the fact that this mutation was observed in the Reunion isolates at an earlier time when compared to Indian isolates, independent introductions in both the countries and subsequent parallel microevolution is most likely.

With regard to global CHIK epidemiological evidence, periodic resurgence of CHIK activity has been reported in different parts of the world interspersed with steady levels of endemicity. Resurgence of CHIK was recorded in 2000 in the DRC (Pastorino et al., 2004), in Indonesia during 2001–2003 (Laras et al., 2005) and in India during 2005–2006 (Yergolkar et al., 2006) after gaps of 39, 20 and 32 years respectively. Though there have been CHIK cases recorded from 1985 to 1999 in Asian countries, overall there seems to be a fairly clear decline in CHIK activity during this 15–20 year period. In the last 2–3 years, there has been an abnormally high increase in CHIK activity in the countries of the Indian Ocean and India. The Bayesian skyline plots obtained in this study, thus provide a notably good reconstruction of the epidemiological data. The pattern may be attributed to the specific amino acid mutations in the virus coupled with modified ecological environments and increasing travel and worldwide movement.

In the present study, the evolutionary timescales of CHIK viruses was estimated to be in the last 300 years with the progenitor of the 2005–07 isolates existing around 9 years ago. The study also confirmed our earlier report (Arankalle et al., 2007) that the progenitor originates from Uganda's neighbourhood. Estimates for the divergence times of different CHIK genotypes has been more accurately determined using the Bayesian MCMC analysis. The present study also implied that for the CHIK viruses, when compared to the whole genome, the E1 gene is adequately suitable for estimation of evolutionary rates and evolutionary time scales through the application of a relaxed molecular clock. The results of the study confirm the epidemiological data with regard to the re-emergence of CHIKV and provide useful information for inferring the plausible direction of spread. The results may further help design strategies for effective disease management.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2008.09.004.

References

Arankalle, V.A., Srivastava, S., Cherian, S., Gunjikar, R.S., Walimbe, A.M., Jadhav, S.M., Sudeep, A.B., Mishra, A.C., 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.

Beltrame, A., Angheben, A., Bisoffi, Z., Monteiro, G., Marocco, S., Calleri, G., Lipani, F., Gobbi, F., Canta, F., Castelli, F., Gulletta, M., Bigoni, S., Del Punta, V., Iacovazzi, T., Romi, R., Nicoletti, L., Ciufolini, M.G., Rorato, G., Negri, C., Viale, P., 2007. Imported Chikungunya infection. Italy. Emerg Infect Dis. 13 (8 (August)), 1264–1266.

Chastel, C., 2005. Chikungunya virus: its recent spread to the southern Indian Ocean and Reunion Island (2005–2006). Bull. Acad. Natl. Med. 189 (8), 1827–1835.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192.

Drummond, A., Ho, J., Phillips, S.Y.W., Rambaut, M.J., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4 (5), 699–710.

Hammon, W.McD., Rudnick, A., Sather, G.E., 1960. Viruses associated with epidemic hemorrhagic fevers of the Philippines and Thailand. Science 131, 1102–1103.

Jobb, G.A., von, Haeseler, Strimmer, K., 2004 Jun 28. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4–18.

Jupp, P.G., McIntosh, B.M., 1988. Chikungunya virus disease. In: Monath, T.P. (Ed.), Arbovirus: Epidemiology and Ecology., Vol. II. CRC Press, Boca Raton, FL, pp. 137–157.

Kokernot, R.H., Szlamp, E.L., Levitt, J., McIntosh, B.M., 1965. Surveys for antibodies against arthropod-borne viruses in the sera of indigenous residents of the Caprivi Atrip and Bechuanaland Protectorate. Trans. R. Soc. Trop. Med. Hyg. 59, 553–652.

Kumar, N.P., Joseph, R., Kamaraj, T., Jambulingam, P., 2008. A226V mutation in virus during the 2007 Chikungunya outbreak in Kerala, India. J. Gen. Virol. 89 (8), 1945–1948.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinf. 5, 150–163.

Lamballerie, X., Leroy, E., Charrel, R.N., Ttsetsarkin, K., Higgs, S., Gould, E.A., 2008. Chikungunya virus adapts to tiger mosquito via evolutionary convergence: a sign of things to come? Virol. J. 5, 33.

Lanciotti, R.S., Kosoy, O.L., Laven, J.J., Panella, A.J., Velez, J.O., Lambert, A.J., Campbell, G.L., 2007. Chikungunya virus in US travelers returning from India, 2006. Emerg. Inf. Dis. 13 (5), 764–767.

Laras, K., Sukri, N.C., Larasati, R.P., Bangs, M.J., Kosim, R., Djauzi, Wandra, T., Master, J., Kosasih, H., Hartati, S., Beckett, C., Sedyaningsih, E.R., Beecham, H.J., Corwin, A.L., 2005 Feb. Tracking the re-emergence of epidemic chikungunya virus in Indonesia. Trans. R. Soc. Trop. Med. Hyg. 99 (2), 128–141.

Macrae, A.W.R., Henderson, B.E., Kirys, B.G., Sempala, S.D.K., 1971. Chikungunya virus in the Entebbe area of Uganda: isolations and epidemiology. Trans. R. Soc. Trop. Med. Hyg. 65, 152–160.

Mourya, D.T., Thakare, J.R., Gokhale, M.D., Powers, A.M., Hundekar, S.L., Jayakumar, P.C., Bondre, V.P., Shouche, Y.S., Padbidri, V.S., 2001. Isolation of chikungunya virus from Aedes aegypti mosquitoes collected in the town of Yawat, Pune District, Maharashtra State. India. Acta Virol. 45 (5–6), 305–309.

National Vector Borne Disease Control Programme (NVBDCP), 2007. Chikungunya fever situation in the country during 2006. <http://www.nvbdcp.gov.in/Doc/CHK.pdf> (Accessed September 11, 2008).

Padbidri, V.S., Gnaneswar, T.T., 1979. Epidemiological investigations of chikungunya epidemic at Barsi, Maharashtra state, India. J. Hyg. Epidemiol. Microbiol. Immunol. 23 (4), 445–451.

Pastorino, B., Muyembe-Tamfum, J.J., Bessaud, M., Tock, F., Tolou, H., Durand, J.P., Peyrefitte, C.N., 2004 Oct. Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain. J Med. Virol. 74 (2), 277–282.

Peyrefitte, C.N., 2007. Chikungunya virus, Cameroon, 2006. Emerg. Infect. Dis. 13 (5 (May)), 768–771.

Peyrefitte, C.N., Bessaud, M., Pastorino, B.A.M., Gravier, P., Plumet, S., Merle, O.L., Moltini, I., Coppin, E., Tock, F., Daries, W., Ollivier, L., Pages, F., Martin, R., Boniface, F., Tolou, H.J., Grandadam, M., 2008. Circulation of Chikungunya virus in Gabon 2006–2007. J Med. Virol. 80, 430–433.

Pialoux, G., Gaüzère, E., Jauréguiberry, S., Strobel, M., 2007. Chikungunya, an epidemic arbovirosis. Lancet Infect Dis. 7 (5 (May)), 319–327.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Powers, A.M., Brault, A.C., Tesh, R.B., Weaver, S.C., 2000. Re-emergence of chikungunya and o'nyong-nyong: evidence for distinct geographical lineages and distinct evolutionary relationships. J. Gen. Virol. 81, 471–479.

Rambaut, A., 2000. Estimating the rate of molecular evolution: incorporating non-comptemporaneous sequences into maximum likelihood phylogenetics. Bioinformatics 16, 395–399.

Rambaut, A., Drummond, A., 2007. Tracer Version 1.4. University of Oxford, Oxford, UK.

Rao, T.R., 1966. Recent epidemics caused by Chikungunya virus in India (1963–65). Sci. Cult. 32, 215–220.

Rodrigues, F.M., Patankar, M.R., Banerjee, K., Bhatt, P.N., Goverdhan, M.K., Pavri, K.M., Vittal, M., 1972. Etiology of the 1965 epidemic of febrile illness in Nagpur city, Maharashtra State, India. Bull. World Health Org. 46 (2), 173–179.

Ross, R.W., 1956. A laboratory technique for studying the insect transmission of animal viruses, employing a bat-wing membrane, demonstrated with two African viruses. J. Hyg (Lond) 54, 192–200.

Santosh, S.R., Dash, P.K., Parida, M.M., Khan, M., Tiwari, M., Lakshmana, Rao., 2008. Comparative full genome analysis revealed E1: A226V shift in 2007 Indian Chikungunya virus isolates. Virus Res 135, 36–41.

Schuffenecker, I., Iteman, I., Michault, A., Murri, S., Frangeul, L., Vaney, M.C., Lavenir, R., Pardigon, N., Reynes, J.M., Pettinelli, F., Biscornet, L., Diancourt, L., Michel, S.,

Duquerroy, S., Guigon, G., Frenkiel, M.P., Bréhin, A.C., Cubito, N., Desprès, P., Kunst, F., Rey, F.A., Zeller, H., Brisse, S., 2006. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 3 (7 (July)), e263 Epub 2006 May 23.

Shah, K.V., Gibbs C.J.Jr., Banerjee, G., 1964 Jul. Virological investigation of the epidemic of haemorrhagic fever in Calcutta: isolation of three strains of Chikungunya virus. Indian J. Med. Res. 52, 676–683.

Strauss, E.G., Strauss, J.H., 1986. Structure and replication of the alphavirus genome. In: Schlesinger, S., Schlesinger, M.J. (Eds.), The Togaviridae and Flaviviridae. Plenum Press, New York, pp. 35–90.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. 18, 1001–1013.

Swofford, D.L., 2003. PAUP*: phylogenetic analysis using parsimony (and other methods), version 4 [computer program]. Sinauer Associates, Sunderland (Massachusetts).

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tsetsarkin, K.A., Vanlandingham, D.L., McGee, C.E., Higgs, S.A., 2007. Single mutation in chikungunya virus affects vector specificity and epidemic potential. Plos Patho 3, e201.

Vazeille, M., Moutailler, S., Coudrier, D., Rousseaux, C., Khun, H., Huerre, M., Thiria, J., Dehecq, J.S., Fontenille, D., Schuffenecker, I., Despres, P., Failloux, A.B., 2007 Nov 14. Two Chikungunya isolates from the outbreak of la reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, *Aedes albopictus*. PLoS ONE 2 (11), e1168.

Weaver, S.C., Rico-Hesse, R., Scott, T.W., 1992. Genetic diversity and slow rates of evolution in New World alphaviruses. Curr. Top. Microbiol. Immunol. 176, 99–117.

Weaver, S.C., Kang, W., Shirako, Y., Rumenapf, T., Strauss, E.G., Strauss, J.H., 1997. Recombinational history and molecular evolution of Western Equine Encephalitis complex alphaviruses. J. Virol. 71 (1), 613–623.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24 (8), 1586–1591.

Yergolkar, P.N., Tandale, B.V., Arankalle, V.A., Sathe, P.S., Sudeep, A.B., Gandhe, S.S., Gokhle, M.D., Jacob, G.P., Hundekar, S.L., Mishra, A.C., 2006 Oct. Chikungunya outbreaks caused by African genotype, India. Emerg. Infect Dis. 12 (10), 1580–1583.