Genome wide analysis of Chikungunya viral strains

Vidhya R. V[1]., Achuthsankar S. Nair[1], Pawan K. Dhar[2]

[1]Department of Computational Biology and Bioinformatics, University of Kerala, Trivandrum 695581, INDIA, Tel (O) 471-2308759.
[2]Centre for Systems and Synthetic Biology, University of Kerala, Trivandrum 695581, INDIA, Tel (O) 471-2308759.

Corresponding Author:
Vidhya R. V., Department of Computational Biology and Bioinformatics, University of Kerala, Trivandrum 695581, INDIA, Tel (O) 471-2308759, Email: amanaparakkat@gmail.com

Abstract

Chikungunya is an ongoing viral epidemic caused by a mosquito–borne viral pathogen, Chikungunya virus (CHIKV). The disease has arisen to prominence during the last decade and affected millions of people around the globe. We know that genetic variations in the viral genomes have important role in imparting virulence to the newly evolved viral strains. In this paper, we have done a genome wide analysis to study the striking regularities and similarities along the genomes of CHIKV strains. The study includes evolutionary analysis, genotype wise- inter and intra group studies and mutation trend analysis. The observations from this study provide new insights into the genetic variations and mutation trends along the genomes of CHIKV genotypes.

Keywords

Phylogenetic analysis, evolutionary calculations, mutation trend, consensus sequence

Introduction

Chikungunya is one of the deleterious ongoing viral epidemics, caused by the viral pathogen, Chikungunya Virus (CHIKV) [1, 2]. The disease is transmitted by *Aedes* genus mosquito vectors and human and non human primates are the primary host of the virus [3]. Person to person transmission has not yet been reported. The disease persists between 2-12 days. Symptoms of the disease are fever, swelling of joints, joint pain, rashes, vomiting, conjunctivae infection, arthralgia, nausea, etc [4, 5]. Till date, there were no specific drugs developed for Chikungunya disease. Chikungunya disease was originated in African countries and expanded to many countries around the globe. The first reported Chikungunya infection was in Tanzania from a febrile patient in 1953[6]. The name, Chikungunya was derived from a Makonde (local language of Makonde Plateau, along the border between Tanganyika and Mozambique) word *Kungunyala*, which means 'that which bends up'. Chikungunya virus, causing Chikungunya infection, is a member of the genus *Alpha* virus in the family *Togaviridae* [7]. Chikungunya virus has a single stranded positive sense non segmented RNA virus, with a genome length of about 11,800 nucleotides [8]. The 5' end of the RNA genome is capped with a 7-methylguanosine and its 3' end is

polyadenylated. The Virus has two reading frames (ORFs), 5'ORF and 3' ORF. The 5'ORF encodes non structural poly protein, which later cleaved into four non structural protein, nsP1-4. A cap dependent mechanism translates structural poly protein from the structural protein ORF, which is included in a subgenomic message and the structural poly protein subsequently cleaved into capsid and E1 and E2 envelop proteins. The genome organization of CHIKV is 5'cap-nsp1-nsP2-nsp3-nsp4-(junction region)-C-E3-E2-6K-E1-poly (A) 3' [9]. Chikungunya virions are spherical shaped with 60-70 nm in diameter having icosahedral nucelocapsid. Previous studies have grouped Chikungunya viral strains mainly into three genotypes; West African, Asian, Eastern, and Central and Southern African (ECSA) [10].

The disease has an infamous history of outbreaks across tropical regions of Africa and Asia. In the last decade, Chikungunya emerged as a dreadful global endemic, emanating severe socioeconomic crisis all over the world. Massive out breaks were reported from African countries, India, the Indian Ocean region, and Southeast Asia. Large scale Chikungunya outbreaks were occurring during that time, in response to the high dissemination rates of the disease, which in turn resulted with considerable diversity in CHIKV as its ensuant effect. Numerous outbreaks were reported from African countries, including Senegal, Comoros, Kenya, Congo, Cameroon, Nigeria, Mayotte, Mozambique, Ghana and, Uganda[11]. The Asian countries affected by the Chikungunya outbreaks were India, Pakistan, Sri Lanka, Vietnam, Thailand, Malaysia, Indonesia, Cambodia, Myanmar, Taiwan, and Philippines. On Indian Ocean islands CHIKV outbreaks have occurred in Mauritius, La Reunion, Seychelles, and Madagascar. Chikungunya cases have been identified from North American countries including Canada and USA and also from Italy [12]. The disease has spread out into various parts of the world and serious Chikungunya cases are being reported even now. The sporadic occurrences of CHIKV outbreaks are due to the genetic variability in Chikungunya viral strains.

Analyzing CHIKV genome sequences and understanding the sequence patterns in Chikungunya viral strains is important in designing suitable preventive measures for controlling the infection. In this paper, a genome wide analysis of CHIKV strains has been conducted to analyze the striking regularities and similarities along the genomes of CHIKV genotypes. The Chikungunya viral sequences that are available in the public domain, that have been used in this work are described below. This paper discusses the groupings of such CHIKV referral sequences and also presents their basic statistical and phylogenetic analysis. Moreover, this study analyzed the mutation trends along CHIKV genomes and subsequently performed a comparative analysis of CHIKV genotypes.

Methods

Selection and Grouping of Referral Sequences

In order to draw knowledge and formulate techniques for identifying genotypes of uncharacterized Chikungunya strains, a Chikungunya referral sequence database has been compiled. Chikungunya sequences from NCBI with their genotypes already identified from

previous studies, as on July 2012, were selected as referral sequences. The referral sequence database consists of a total of 78 complete genomic sequences of Chikungunya virus classified into five groups as given in Table 1[11,13]. For easy reference, brief IDs of referral sequences and their corresponding NCBI IDs have been given in Appendix 1.

Sl. No	CHIKV Genotype	Reference Sequence	No. of Sequences
1	Group I	R1-R12	12
2	Group II	R13-R34	22
3	Group IIIa	R35-R41	7
4	Group IIIb	R42-R46	5
5	Group IIIc	R47-R78	32

Table 1: Summary table of CHIKV Referral sequences
(See Appendix 1 for NCBI IDs)

Chikungunya is presently classified into three major genotypes, Group I, II and III. Group I is the West African Genotype, Group II is the Asian Genotype, and Group III is East /Central/ South African Genotype (ECSA). Group III is further divided into three subgroups: East/South/Central African subgroup (IIIa), Indian Ocean sub group (IIIb) and Asian subgroup (IIIc) [11, 13]. The country of origin and year of reporting of each of the group are shown in Table 2[10, 14].

Genotype	Subtype	Country of Origin	Years of Reporting
Group I		Senegal, Nigeria, Cote d'Ivoire	1963-2005
Group II		India, Indonesia, Thailand, Malaysia, USA, Philippines	1958-2007
Group III	IIIa	Tanzania, South Africa, Democratic Republic of the Congo, India	1953-1986
	IIIb	Angola, Central African Republic, Uganda	1962-1982
	IIIc	Mauritius, Reunion, India, Sri Lanka, Singapore, Italy	2006-2012

Table 2: Details of Country of origin and years of reporting of CHIKV Referral Sequences

Phylogenetic Analysis

In order to confirm the groupings in CHIKV, a phylogenetic tree has been constructed using our referral strains. Phylogenetic tree reconstructs the evolution of life and it is usually performed to infer evolutionary relationship between organisms under consideration [15]. Phylogenetics try to cluster living things based on their level of similarity and it illustrates the divergence pattern through the tree topology [16]. Molecular phylogenetics is a branch of phylogenetics that deals with molecular level data like genomic sequences, DNA, RNA, or protein sequences. The basic concept of phylogenetics is that the members of a clade in the phylogenetic tree share a common evolutionary history and these are more related to each other than to members of another group. Computational tree building tools apply statistical approaches for generating phylogenetic trees. Most popular tree building packages are MEGA, PHYLIP, PAUP and PAML [17]. Here, a neighbor-joining (NJ) tree of CHIKV was

generated, using MEGA 5 tool, and a boot strap analysis was performed for testing the reliability of clades within the phylogenetic tree with 1000 replication cycles.

Preliminary group-wise analysis based on pair wise distances

In this step, pair wise distances between genomic sequences in each of the CHIKV genotypes have been computed. Pair wise distance is a measure of character differences between pair of sequences and can provide reasonable estimate of phylogenetic relationship [18]. These distance values explain the difference or similarity of each pair of taxa. Here, the pair wise distance estimation was performed using MEGA 6 [19] and mean and standard deviation of these pair wise distances in CHIKV genotypes were also estimated.

Distance Analysis based on Consensus Sequence

For analyzing the variations within CHIKV groups, we also took an alternate approach. We computed the consensus sequences for each of the CHIKV groups using EMBL Consensus tool [20]. Consensus sequence is a theoretical concept and it is generated from multiple sequence alignment. Consensus sequence is generated by identifying and displaying the most frequent residues at each position in the multiple sequence alignment [21]. It helps to identify the conserved and variable regions along the sequences and to understand the relatedness of sequences. In order to check the reliability of the tool, consensus sequences were created using another consensus generating tool, EMBOSS cons [22]. After that, pair wise distances and similarity percentages between consensus sequences of CHIKV genotypes have been estimated to identify similarities and distances among CHIKV genotypes. Further, an inter-group phylogenetic tree was also constructed to analyze relatedness between CHIKV genotypes.

Evolutionary Calculations

Calculating evolutionary parameters provide information regarding sequence evolution. Currently available phylogenic programs have inbuilt tools for estimating these evolutionary parameters [19].

Evolutionary divergence

This parameter estimates the evolutionary divergence among sequences under consideration. For calculating sequence divergence, mean distance of all the sequences to a reference sequence is estimated and then the diversity is measured as the mean distance between all sequences. Sequence divergence can be calculated using the formula [23].

$$D_{diversity} = \frac{1}{N(N-1)} \sum_{\substack{i,j \\ i \neq j}}^{N} d(i,j)$$

Where N is the total number of sequences, d (i, j) is the genetic distance between sequence i and j. The Tajimas D for CHIKV genotypes have been estimated using MEGA tool [19].

Gamma parameter

Substitution rates at sequence positions along the sequence may not be constant; it varies from sites to sites. This rate distribution is calculated by a constant known as Gamma parameter. The higher gamma parameter values indicate lower substitution rates and lower gamma parameter value correspond to higher substitution rates. Gamma distribution parameter for CHIKV genotypes have been computed using MEGA 6 tool [19].

Tajimas statistical test

Tajimas D statistics is a commonly used neutrality test for analyzing the evidence of deviation from standard neutral model. This tests the effect of natural selection on the sequence dataset. Tajimas D can be calculated as follows;

$$D = \frac{\widehat{\theta}_\pi - \widehat{\theta}_S}{\sqrt{\widehat{\mathbf{Var}}[\widehat{\theta}_\pi - \widehat{\theta}_S]}}.$$

The Tajimas D value is determined by two θ estimators; Tajima's estimator (θ_π) and Watterson's estimator (θ_s). Tajima's estimator is based on the average number of pair wise differences between sequences in the dataset and the Watterson's estimator based on the number of segregating sites in the sequences. A negative value of Tajima's D indicates population size expansion due to an excess of rare polymorphisms relative to expectation. A positive Tajima's D values show the presence of high frequency polymorphisms that leads to decrease in population size indicating balancing selection [19].

Analysis of Mutation trends along CHIKV genome

Mutation frequencies have been noted from multiple sequence alignments, in order to study mutation trends along CHIKV genome. At first, mutation trends along each of the CHIKV genotypes were analyzed by plotting mutation frequency Vs nucleotide length graph. Subsequently, consensus sequences from all the CHIKV genotypes were observed for plotting mutation trends along CHIKV genome. Followed by this all the referral sequences in CHIKV dataset have been considered for mutation trend analysis. Further, the mutation trends in genomic regions that code for structural and non-structural poly proteins were examined. Using these analysis results, i.e. from mutation frequency Vs nucleotide length graph, the frequently mutating regions along the CHIKV genome were recognized and the proteins corresponding to these mutational hotspot regions were identified.

Results

Selection and Grouping of Referral Sequences

A CHIKV referral dataset has been generated by selecting all the available CHIKV genomic sequences from NCBI database. Based on the information available from the literature and database the retrieved CHIKV referral sequences were grouped into five groups. A phylogenetic analysis was also performed to confirm the groupings. The tree is as shown in Figure 1. In order to confirm the grouping obtained from NJ tree, other phylogenetic trees

have been generated from the same dataset, using Minimum Evolution, UPGMA (Unweighed Pair Group Method with Arithmetic Mean), Maximum Parsimony and Maximum likelihood methods. Majority of the methods gave the same classification as in Figure 1, increasing its confidence.

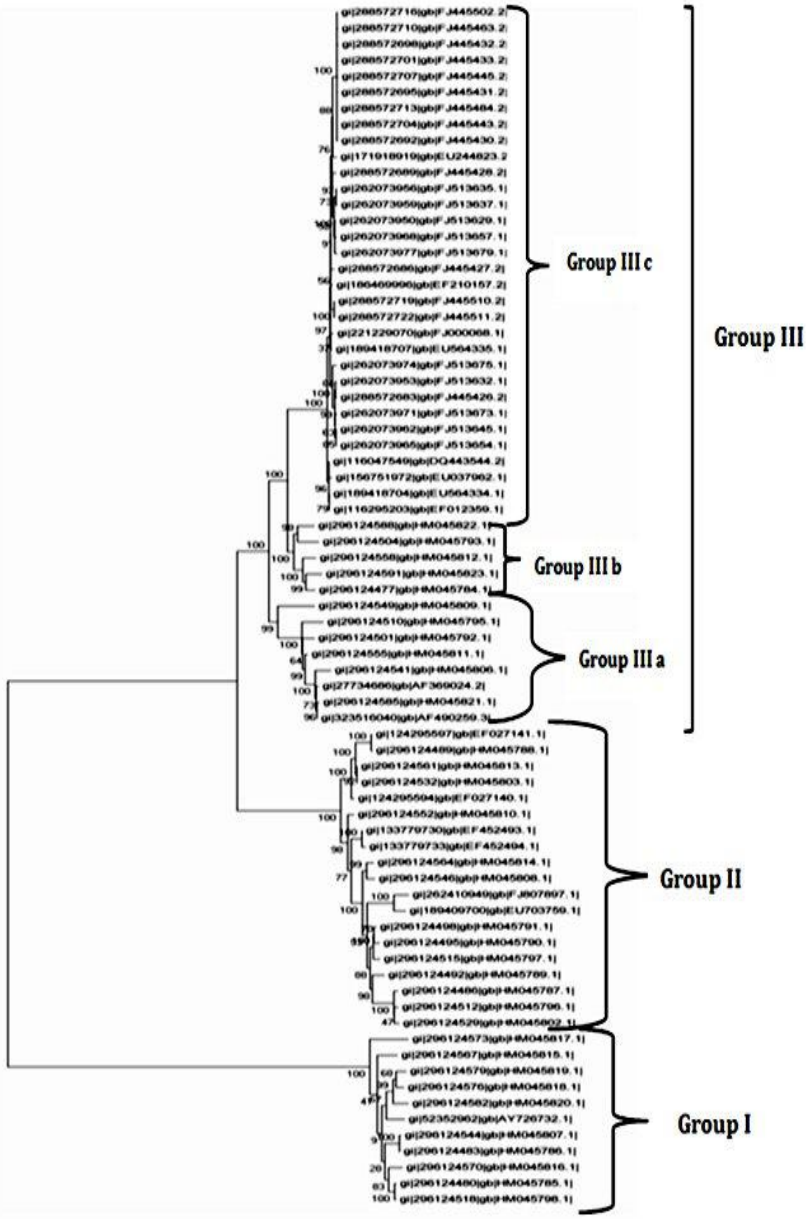

Figure 1: Phylogenetic tree of CHIKV referral strains

The observed phylogenetic tree double confirms the grouping in CHIKV. From the tree, it is evident that there are three major clusters that correspond to Group I, II and III. Group III has three sub clusters and these clusters represent Group IIIa, IIIb and IIIc. It is also

observed that the other groups were evolved from Group I, West African strains, as this group forms the out group of the tree.

Preliminary group wise analysis based on pair wise distances

Pair wise distance between all the Group I strains were computed using the phylogenetic software MEGA 6 based on maximum composite likelihood substitution model. The results are given in Table 3, along with the mean and standard deviation of these pair wise distances. The mean and standard deviation of group I are 0.009 and 0.005 respectively.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
R1	0.000											
R2	0.000	0.000										
R3	0.000	0.008	0.000									
R4	0.008	0.008	0.007	0.000								
R5	0.007	0.008	0.005	0.005	0.000							
R6	0.007	0.008	0.001	0.005	0.010	0.000						
R7	0.010	0.010	0.006	0.008	0.005	0.017	0.000					
R8	0.008	0.008	0.010	0.011	0.010	0.013	0.016	0.000				
R9	0.008	0.009	0.011	0.011	0.011	0.014	0.007	0.008	0.000			
R10	0.010	0.010	0.013	0.013	0.013	0.016	0.004	0.010	0.012	0.000		
R11	0.010	0.011	0.013	0.013	0.013	0.016	0.005	0.010	0.008	0.008	0.000	
R12	0.014	0.014	0.017	0.017	0.017	0.019	0.010	0.014	0.012	0.013	0.012	0.000

Table 3: Pair wise distance analysis of Group I, Mean=0.009; Standard deviation=0.005

Similarly, pair wise distances, mean and standard deviation of all the CHIKV genotypes have been calculated. The mean and standard deviation of the preliminary pair wise distance analysis of Group I, II and Group III (a-c) are given in Table 4.

CHIKV group	Mean	Standard Deviatio
Group I	0.009	0.005
Group II	0.012	0.007
Group IIIa	0.008	0.007
Group IIIb	0.006	0.005
Group IIIc	0.002	0.001

Table 4: Mean & Standard deviation of CHIKV genotypes

Distance Analysis based on Consensus Sequence

The consensus sequence of the Group-I referral sequences were generated using EMBL Consensus tool. Threshold parameter was selected as 90. The consensus sequence for Group I is given in Figure 2.

>west African
GTAGCCTACCAGTTTCTTACTGCTCTACTCTGCTTAGCAAGAGACTTGAGAACCCATCATGGATCCCGTGT
ACGTGGACATAGACGCCGACAGCGCCTTTTTGAAGGCCCTGCAGCGTGCGTACCCCATGTTTGAGGTGGAA
CCAAGGCAGGTCACACCGAATGACCATGCCAATGCTAGAGCATTCTCGCATCTAGCTATAAAACTAATAGA
GCAGGAAATTGATCCCGACTCAACCATCCTGGACATAGG.....

Figure 2: Consensus Sequence of Group I Referral Sequences: First 250 bases

The distance analysis and phylogenetic analysis performed above have captured the intra-group relations between sequences in each group. To get a clear picture of the uniqueness of groups, the pair wise distances and similarity percentages between consensus sequences of each of the CHIKV genotype groups have also been estimated. The inter group pair wise distances and similarity percentages are shown in Table 5 and 6 respectively.

	Group I	Group II	Group IIIa	Group IIIb	Group IIIc
Group I	0				
Group II	0.047	0			
Group IIIa	0.042	0.013	0		
Group IIIb	0.042	0.013	0.019	0	
Group IIIc	0.045	0.015	0.007	0.004	0

Table 5: Inter group pair wise distances of CHIKV consensus sequence

	Group I	Group I	Group III	Group III	Group IIIc
Group I	0				
Group II	85%	0			
Group IIIa	86%	95%	0		
Group IIIb	86%	95%	98%	0	
Group IIIc	85%	98%	97%	98%	0

Table 6: Intergroup pair wise similarity percentages of CHIKV consensus sequences

Finally, an inter-group phylogenetic tree based on consensus sequences of all CHIKV genotypes has been constructed to verify the previous observations. The inter group phylogentic tree is displayed in Figure 3. This also confirms that the other CHIKV groups were evolved from Group I West African strains.

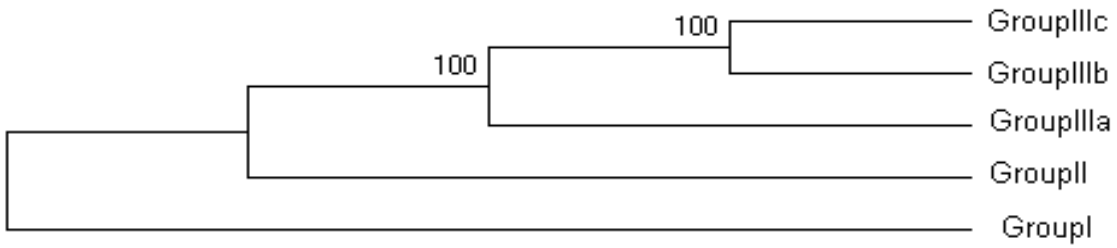

Figure 3: Inter-group phylogenetic tree of CHIKV genotype groups

A genome wide analysis of each of the CHIKV genotypes has been conducted to enumerate variable sites along the genome. Percentages of variable regions are listed in Table 7.

Chikungunya Genotype	Number of variable sites along the genome	Variable sites (%)
Group I	446	3.81%
Group II	628	5%
Group IIIa	376	3%
Group IIIb	252	2%
Group IIIc	191	2%

Table 7: Percentages of variable regions along the genomes of CHIKV genotypes

Group II genotype is having a higher percentage of variable regions compared to other genotypes. So, we can conclude Group II genotypes diverge much faster than other genotypes.

Evolutionary Calculations

Evolutionary statistical parameters were computed for analyzing molecular evolution of strains in each of the CHIKV genotypes. In this section, three evolutionary parameters; Evolutionary divergence, Gamma distribution parameter and Tajimas statistical test have been estimated. The results are displayed in Table 8-10.

CHIKV Genotype	Evolutionary Divergence
Group I	0.010
Group II	0.014
Group IIIa	0.010
Group IIIb	0.010
Group IIIc	0.002

Table 8: Evolutionary divergence of CHIKV genotypes

Evolutionary divergence values indicate that Group II is a less conserved group with high evolutionary divergence value. The lower evolutionary divergence value of Group III suggests that evolution in Group IIIc has reached a steady state.

CHIKV group	Gamma parameter
Group I	0.2492
Group II	0.3654
Group IIIa	0.4408
Group IIIb	0.0975
Group IIIc	0.0500

Table 9: Gamma distribution parameters of CHIKV genotypes

Gamma distribution parameter values of CHIKV genotypes show that the highest substitution rates among CHIKV groups is observed in Group IIIa strains.

CHIKV genotype	m	S	Ps	ɵ	Π	D
Group I	12	447	0.0377	0.0129	0.0101	-1.0425
Group II	22	625	0.0537	0.0147	0.0134	-0.3621
Group IIIa	7	358	0.0310	0.0126	0.0099	-1.2667
Group IIIb	5	251	0.0214	0.0103	0.0095	-0.55621
Group IIIc	32	174	0.0150	0.0037	0.0024	-1.40334

Table 10: Tajimas D values obtained for CHIKV genotypes

For all the CHIKV genotypes, the calculated Tajimas D values were observed as negative values hence the null hypothesis of mutation-drift equilibrium and constant population size got rejected for CHIKV genotypes. Tajima's D test shows excess of low frequency polymorphisms in CHIKV Groups relative to expectation, indicating population size expansion.

Analysis of Mutation trends along CHIKV genome

Mutation trends were analyzed by plotting mutation frequency graphs, for this, nucleotide lengths were plotted along X-axis and mutation frequencies were plotted along Y axis. First, mutation frequency graphs for all the CHIKV genotypes were generated. The obtained graphs were displayed in Figure 4a-e.

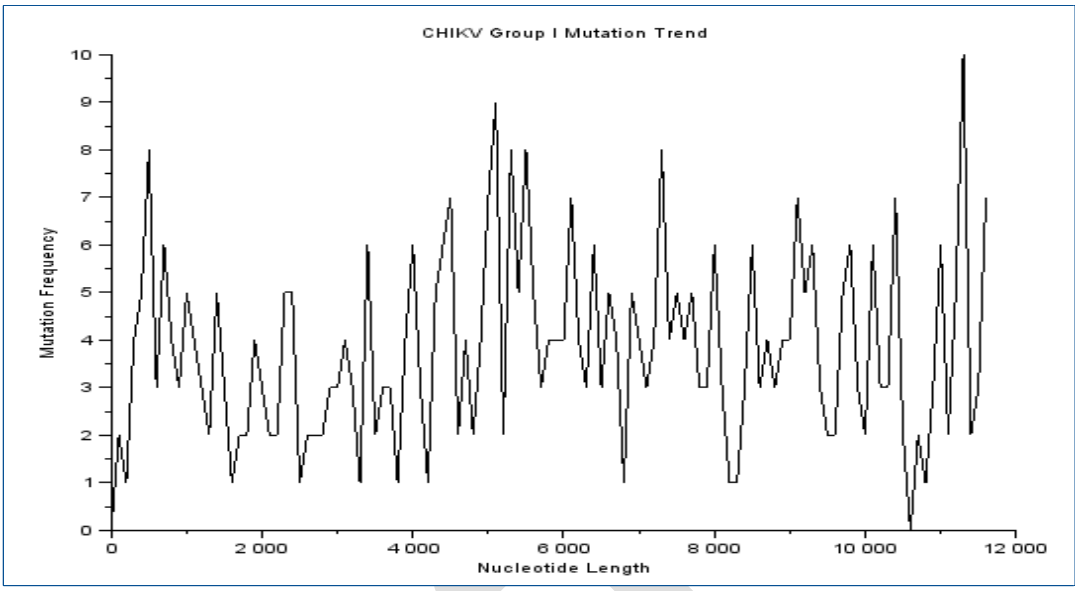

Figure 4a: Mutation analysis of Group I strains

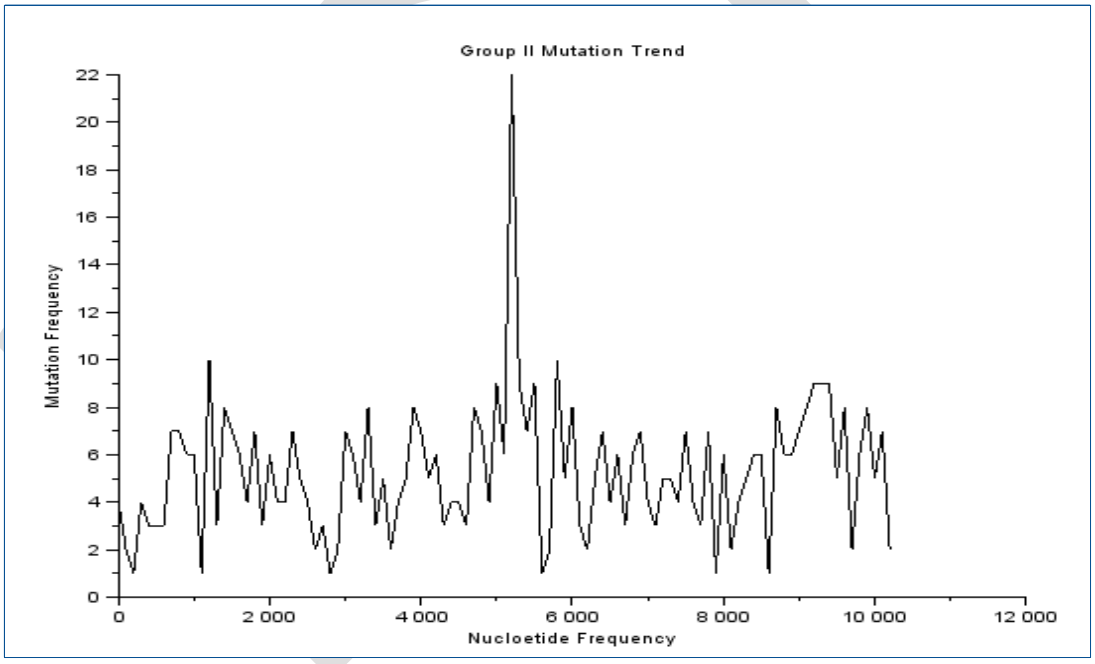

Figure 4b: Mutation analysis of Group II strains

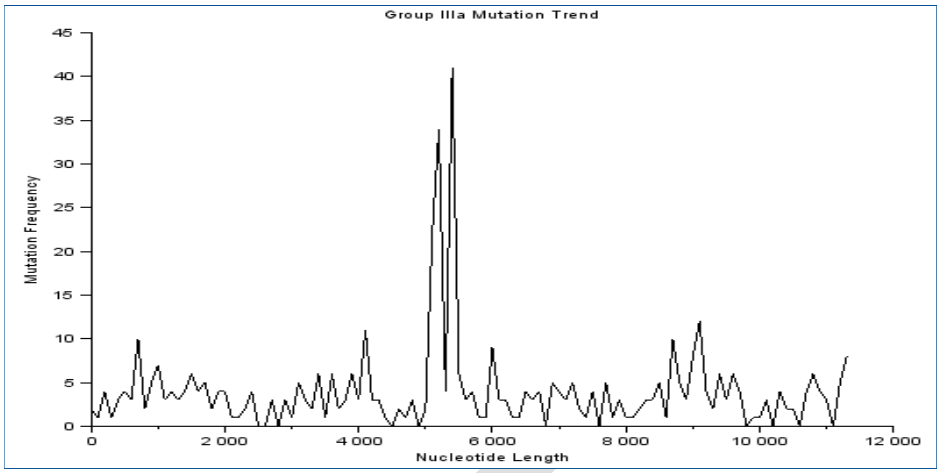

Figure 4c: Mutation analysis of Group IIIa strains

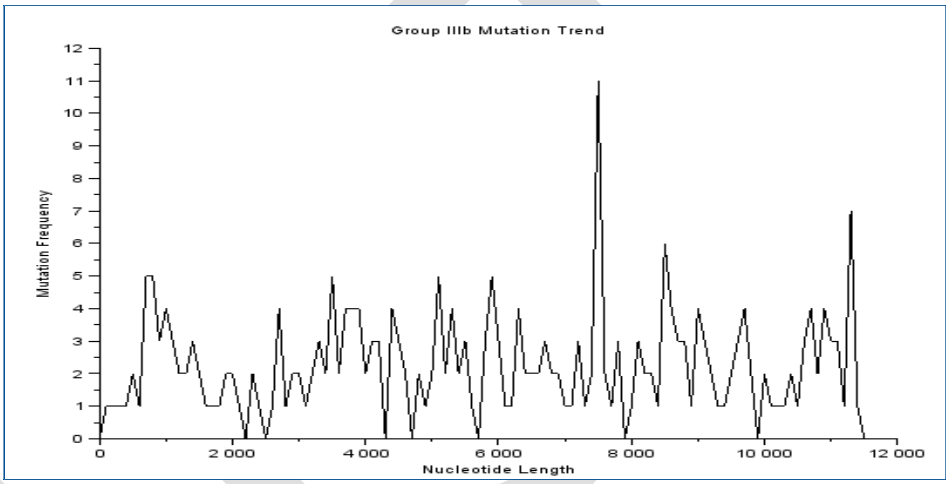

Figure 4d: Mutation analysis of Group IIIb strains

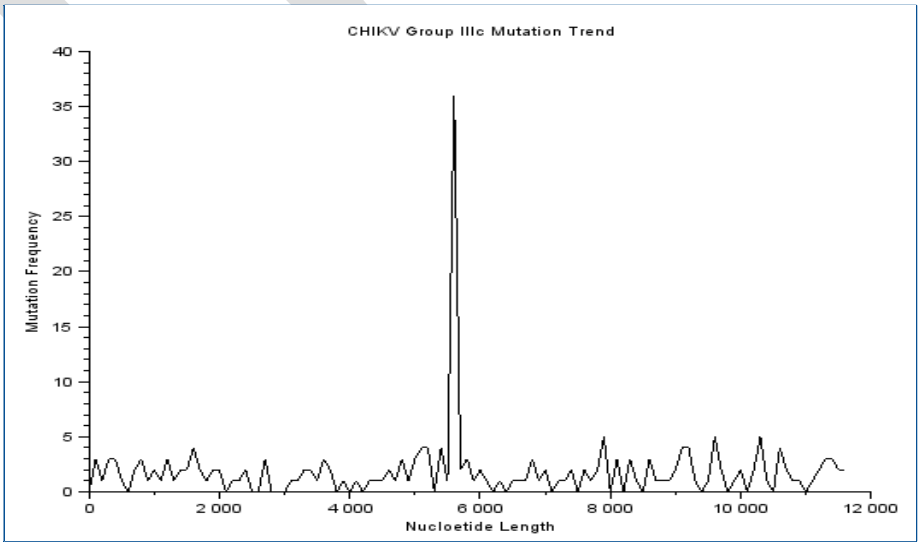

Figure 4e: Mutation analysis of Group IIIc strains

Consensus sequences mutation frequency graphs were also generated using the consensus sequences from CHIKV genotypes. The graph showing the mutation trends is shown in Figure 5.

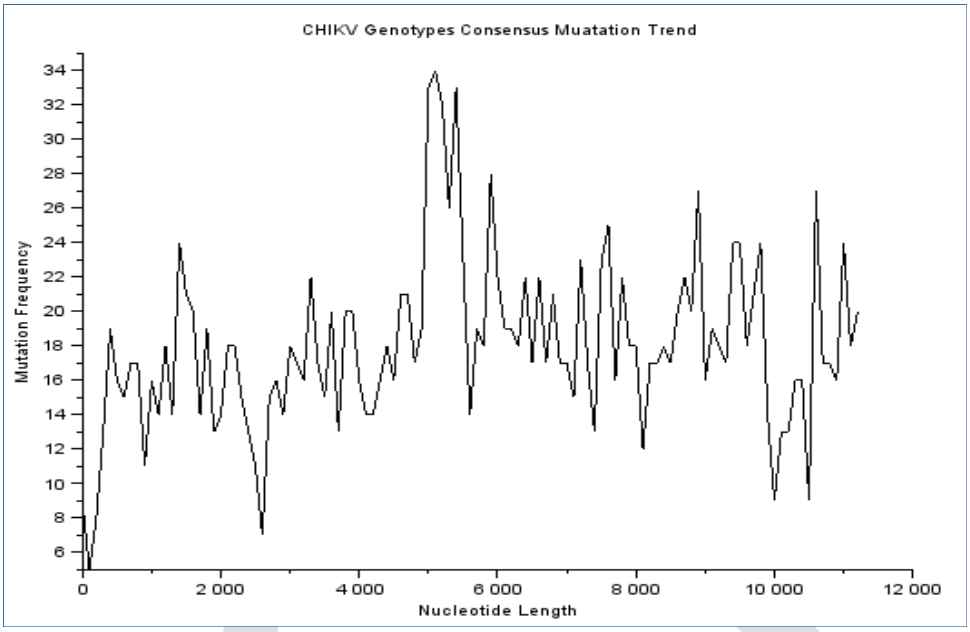

Figure 5: Mutation analysis of CHIKV consensus sequences

Followed by this, mutation frequency graph using all the CHIKV referral sequence has been plotted, and the obtained graph is displayed in Figure 6. Further, mutation trends in genomic regions that code for structural and non-structural poly proteins were also analyzed.

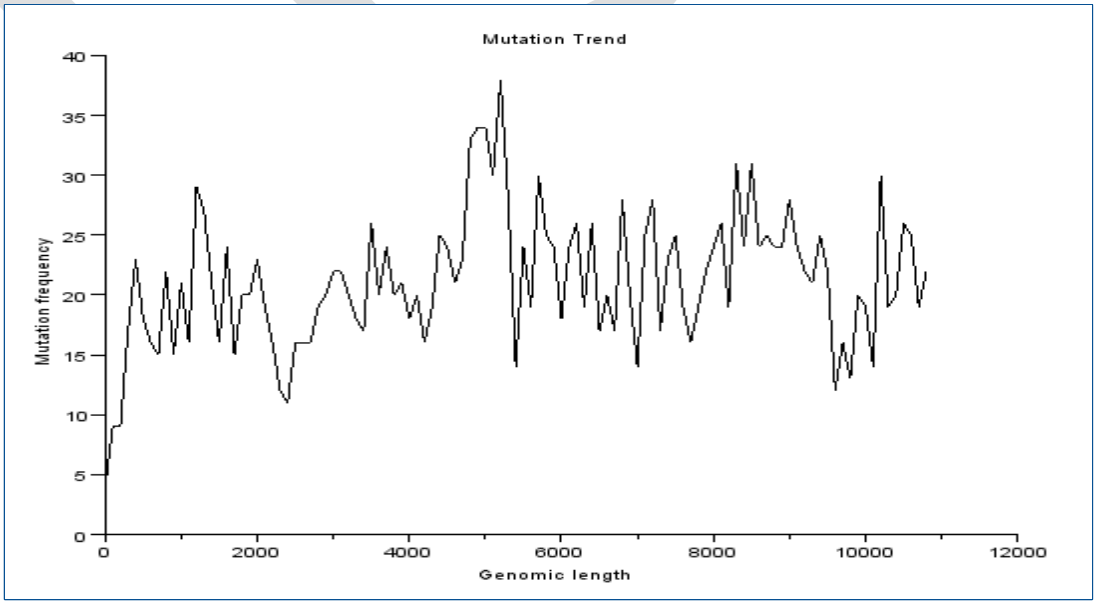

Figure 6: Mutation analysis of CHIKV referral sequences

Peaks observed in this graphs correspond to regions having highest mutation frequencies. From these mutation frequency graph it was identified that the genomic regions (4500 – 6000) & (8400-10000) are the mutational hot spots along the CHIKV genome. The proteins corresponding to these frequently mutating regions are nsp3-nsP4, and E3–E2[24].

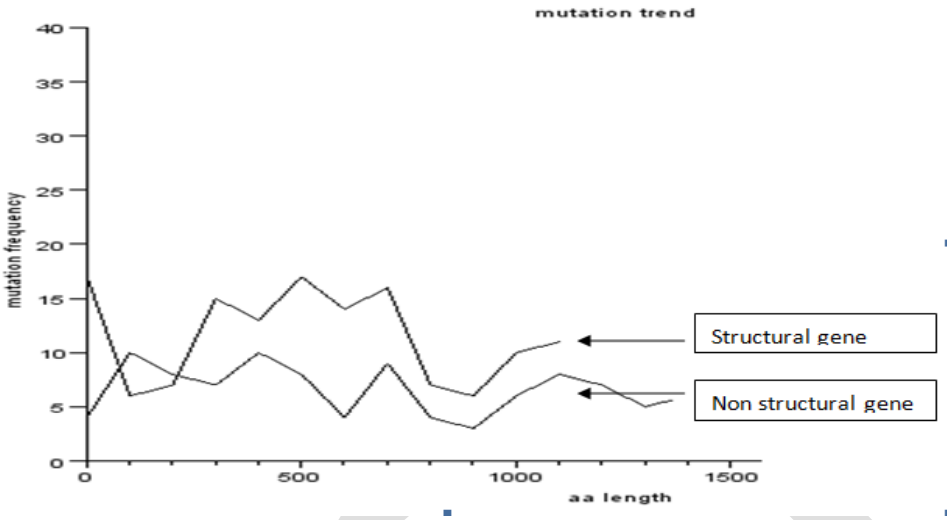

Figure 7: Mutation analysis of CHIKV genomic region that codes for structural and non-structural poly proteins

Analysis of mutation trends in CHIKV structural & non structural poly protein regions shows the structural poly protein mutate more rapidly compared to non structural poly protein. The graph is displayed in Figure 7.

Conclusion

In this study, a CHIKV referral dataset have been created and used these referral sequences for preliminary genome wide analysis. The phylogenetic tree obtained in this study double confirmed the grouping in CHIKV strains. Based on the resultant tree and also using the information retrieved from literature and sequence databases, CHIKV referral strains were grouped into five separate groups; Group I consist of west African strains, Group II include Asian Strains and Group III consist of East Central South African strains. Group III has three subgroups; Group IIIa, IIIb and IIIc. In order to derive some quantitative measures of the referral sequences, pair wise distance analysis and evolutionary calculations on CHIKV groups have been performed. Pair wise distance calculations suggest that CHIKV group IIIc has reached its steady state of evolution and the Asian genotype (Group II) can be said as an evolutionary active, evolving group. Consensus sequence based pair wise calculations were also done on these CHIKV genotypes to investigate the intergroup relationship between CHIKV genotypes and result obtained from this analysis suggests that other CHIKV groups were evolved from Group I strains. The estimated evolutionary divergence values explain the degree of variance among CHIKV strains with in each of the CHIKV genotype. In this chapter, rate variations among sites in CHIKV genotypes were also

estimated and the lowest variation of substitution rates among sites was observed in group IIIa strains. For examining the homologous sequence frequency distribution, Tajimas D parameter has been calculted. For all the CHIKV genotypes, the analysis got negative values for Tajimas D indicating population size expansion in all the CHIKV genotypes. In addition to that, the Tajima D value obtained for Group II strains suggests excess of genetic variations in CHIKV Group II. This result support observation from pair wise analysis that CHIKV is an evolving group with excess of genetic variations. Mutation trend graphs resulted from conservation variation analysis of CHIKV strains identified mutation hot spots (4500 – 6000) and (8400-10000) along the CHIKV genome and it was found that the protein corresponding to the frequently mutating regions are nsp3-nsP4, and E3–E2. Analysis of mutation trends in CHIKV structural & non-structural poly protein genomic regions revealed that the structural poly protein mutate more frequently than non-structural poly protein region. The preliminary analysis described in this chapter provides a better understanding of genome based aspects about the CHIKV strains.

Acknowledgment

The corresponding author acknowledges the funding from Kerala State IT Mission, Govt. of Kerala for the SPEED- IT (Special Postgraduate Education Expansion Drive in IT) Fellowship.

References

1. Pialoux G, Gauzere BA, Jaureguiberry S, Strobel M (2007) Chikungunya, an epidemic arbovirosis. Lancet Infect Dis 7: 319–327
2. Gubler DJ (2001) Human arbovirus infections worldwide. *Ann N Y Acad Sci* 951: 13–24.
3. Sunit Kumar Singh and Salini Krishnan Unni (2011) Chikungunya virus: host pathogen interaction, Reviews in Medical Virology 21(2):78–88.
4. Ananthanarayan R, Paniker CKJ (1986) Textbook of microbiology, 3rd ed. Madras: Orient Longman Ltd.: 502-7.
5. Ozden S, Huerre M, Riviere JP, Coffey LL, Afonso PV, Mouly V, et al.(2007) Human muscle satellite cells as targets of chikungunya virus infection. PLoS One; 2: e527.
6. Robinson MC (1955) An epidemic of virus disease in southern province, Tanganyika territory, in 1952-53, Trans R Soc Trop Med Hyg., 49(1): 28–32.
7. Das S , Kolhe RP, Mane BG, Singh JP Nand Singh AP (2007) Chikungunya Epidemic: Global and Indian scenario, J. Commun. Dis., 39 (1): 37-43.
8. Adel A. Rashad, Paul A. Kelle (2013) Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins, Journal of Molecular Graphics and Modelling, 44 : 241-252.
9. Scott C Weaver et. al (2012) Chikungunya Virus and Prospects for a Vaccine, Expert Rev Vaccines, 11(9) :1087-1101.
10. A.B. Sudeep, D. Parashar (2008) Chikungunya: an overview, Journal of Biosciences, 33(4): 443-449.
11. Volk SM, Chen R, Tsetsarkin KA, Paige Adams A, 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(13) :6497–6504. doi:10.1128/JVI.01603-09.

12. Francesca Cavrini, Paolo Gaibani, Anna Maria Pierro, Giada Rossini, Maria Paola Landini and Vittorio Sambri (2009), Chikungunya: an emerging and spreading arthropod-borne viral disease, J Infect Dev Ctries,3(10):744-752.
13. De DW, Yonghui Z, Qiong Z, Jing K, Wenjia L, Huan Z, Corina M, Qiaoli Z, Wenjie L, Haojie Z, Jianfeng H, Hui L, Songwu C, Changwen K, Jinyan L (2013) Chikungunya virus with E1- A226V mutation causing two outbreaks in 2010, Guangdong, China, Virol J, 10:174. doi:10.1186/1743-422X-10-174.
14. Cui J Gao M, Ren X (2011) Phylogeny and homologous recombination in Chikungunya viruses, Infect Genet Evol., 11(8):1957-63. doi:10.1016/j.meegid.2011.08.026.
15. Bernard M. E. Moret, David A. Bader, Tandy Warnow (2001) High-Performance Algorithmic Engineering for Computational Phylogenetics, Computational Science, 2074: 1012-1021.
16. Donald R. Forsdyke (2006) Evolutionary Bioinformatics, Springer, New York Dordrecht Heidelberg London.
17. Durbin R, S Eddy, A Krogh, G Mitchison (1998) Biological Sequence Analysis. Cambridge University Press: Cambridge.
18. Yves Van de Peer (2009) The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing, Philippe Lemey, Marco Salemi, and Anne-Mieke Vandamme (eds.). Published by Cambridge University Press.
19. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol Biol Evol.: doi:10.1093/molbev/msr121.
20. http://coot.embl.de/Alignment/consensus.html
21. John C. Obenauer, Jackie Denson, Perdeep K. Mehta, Xiaoping Su, Suraj Mukatira, David B. Finkelstein, Xiequn Xu, Jinhua Wang, Jing Ma, Yiping Fan, Karen M. Rakestraw, Robert G. Webster, Erich Hoffmann, Scott Krauss, Jie Zheng, Ziwei Zhang, Clayton W. Naeve (2006) Large-Scale Sequence Analysis of Avian Influenza Isolates, Science, 311(5767):1576-1580.
22. Mullan LJ, Bleasby AJ (2002) Short EMBOSS user guide, Eur Mol Biol Open Softw Suite Brief Bioinform 3(1):92–94.
23. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J. Virol., 73:10489–10502.
24. Maxime Solignat, Bernard Gay, Stephen Higgs, Laurence Briant, Christian Devaux (2009) Replication cycle of chikungunya: A re-emerging arbovirus, Virology, 393(2): 183–197.

Appendix 1

Brief IDs of Referral sequences and corresponding NCBI IDs

Table A1.1: Details Group I (West African Strains) CHIKV Referral Sequences

Reference No	NCBI ID	Country
R1	HM045786	Nigeria
R2	HM045807	Nigeria
R3	HM045816	Senegal
R4	HM045798	Senegal
R5	HM045785	Senegal
R6	HM045815	Senegal
R7	HM045818	Cote d'Ivoire
R8	AY726732	Senegal
R9	HM045820	Cote d'Ivoire
R10	HM045819	Senegal
R11	HM045817	Senegal
R12	HM045804	Senegal

Table A1.2: Details Group II (Asian Strains) in the Reference Database

Reference No	NCBI ID	Country
R13	HM045810	Thailand
R14	EF452493	Thailand
R15	EF027140	India
R16	HM045803	India
R17	HM045813	India
R18	EF027141	India
R19	HM045788	India
R20	HM045814	Thailand
R21	HM045808	Thailand
R22	HM045791	Indonesia
R23	HM045797	Indonesia
R24	HM045790	Philippines
R25	HM045789	Thailand
R26	HM045796	Thailand
R27	HM045787	Thailand
R28	EU703759	Malaysia: BaganPanchor
R29	EU703761	Malaysia: BaganPanchor
R30	EU703762	Malaysia: BaganPanchor
R31	EU703760	Malaysia: BaganPanchor
R32	FJ807897	Indonesia
R33	HM045802	Thailand
R34	EF452494	USA

Table A1.3: Details Group III a CHIKV Strains in the Reference Database

Reference No	NCBI ID	Country
R35	AF 369024	African
R36	HM045811	Tanzania
R37	HM045792	South Africa
R38	AF490259	unknown
R39	HM045809	Democratic Republic of the Congo
R40	HM045795	South Africa
R41	HM045806	India

Table A1.4: Details Group III b CHIKV Strains in the Reference Database

Reference No	NCBI ID	Country
R42	HM045823	Angola
R43	HM045822	Central African Republic
R44	HM045784	Central African Republic
R45	HM045793	Central African Republic
R46	HM045812	Uganda

Table A1.5: Details Group IIIc CHIKV Strains in the Reference Database

Reference No	NCBI ID	Country
R47	DQ443544	Reunion
R48	EU037962	Mauritius
R49	EU564334	Mauritius
R50	EF012359	Mauritius
R51	FJ000068	India- Karnataka
R52	EF210157	India
R53	EU564335	India- Rajasthan
R54	FJ513675	Sri Lanka
R55	FJ513657	Sri Lanka
R56	FJ513645	Sri Lanka
R57	FJ513635	Sri Lanka
R58	FJ513629	Sri Lanka
R59	FJ513679	Sri Lanka

R60	FJ513673	Sri Lanka
R61	FJ513654	Sri Lanka
R62	FJ513637	Sri Lanka
R63	FJ513632	Sri Lanka
R64	FJ445428	Sri Lanka
R65	FJ445426	Sri Lanka
R66	FJ445427	Sri Lanka
R67	FJ445510	Singapore
R68	FJ445484	Singapore
R69	FJ445445	Singapore
R70	FJ445433	Singapore
R71	FJ445431	Singapore
R72	FJ445511	Singapore
R73	FJ445502	Singapore
R74	FJ445463	Singapore
R75	FJ445443	Singapore
R76	FJ445432	Singapore
R77	FJ445430	Singapore
R78	EU244823	Italy