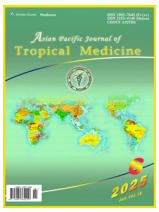


Original Article

Asian Pacific Journal of Tropical Medicine

apjtm.org

doi: 10.4103/apjtm.apjtm_746_24

Exploring the impact of envelope protein mutations on Chikungunya virus epitopes: Analysis of virus samples from the Alagoas State outbreak, Brazil

Jamile Taniele-Silva[1], Júlia De Andrade Brandão[1], Maria Júlia Tenório Costa Cinésio De Oliveira[1], Stephannie Janaina Maia De Souza[1], Jean Fábio Gomes Ferro[2], Magliones Carneiro De Lima[2], Abelardo Silva-Júnior[1], Ênio José Bassi[1]✉, Letícia Anderson[3]✉

[1]*Laboratório de Pesquisas em Virologia e Imunologia (LAPEVI), Instituto de Ciências Biológicas e da Saúde (ICBS), Universidade Federal de Alagoas, 57072-900, Maceió, Alagoas, Brazil*

[2]*Laboratório Central de Saúde Pública de Alagoas (LACEN/AL), 57036-860, Maceió, Alagoas, Brazil*

[3]*Laboratório de Expressão Gênica e Estudos Moleculares (LEGEM), Instituto de Química e Biotecnologia (IQB), Universidade Federal de Alagoas, 57072-900, Maceió, Alagoas, Brazil*

ABSTRACT

Objective: To investigate mutations in the Chikungunya (CHIKV) envelope genome region and evaluate their potential impact on B lymphocyte epitopes *via in silico* analysis.

Methods: E1, E2 and 6K protein genes were sequenced from viral RNA isolated from 13 CHIKV-positive serum samples from Alagoas State, Brazil, during the 2016 outbreak. Phylogenetic analysis, experimental epitope identification in the immune epitope database (IEDB) and *in silico* approaches were employed to predict the potential impact of the detected mutations.

Results: The sequences were clustered *via* phylogenetic analysis. The CHIKV isolates belong to the ECSA genotype, with 13 detected amino acid mutations. Five mutations are located on the surface of the viral particle in regions critical for cellular receptor interaction. Nine mutations are known experimentally validated epitopes for B and T cells. In B-cell epitope predictions, mutations affect sequences within three conformational epitopes in E2 and one in E1, as well as linear epitopes. Notably, the E2-G60D mutation found in the Alagoas strain has been previously reported to influence the vector competence of *Aedes aegypti*, the primary vector in Brazil.

Conclusions: Genomic surveillance and an in-depth understanding of viral mutations are crucial for adapting public health strategies and improving the outbreak response. These findings could have significant public health implications, such as the development of more effective vaccines, diagnostic tests, and antiviral therapies.

KEYWORDS: Genomic variations; *In silico* analysis; Epitope prediction; Glycoprotein mutations

1. Introduction

Chikungunya virus (CHIKV) is a re-emerging arbovirus with a high global impact due to its increasing spread on new continents, which has been attributed to several extrinsic factors, such as climate change, increased urbanization, high population density, and intrinsic factors, such as the ability of the virus to adapt genetically through mutations in its genome[1]. CHIKV was reported for the

Summary

Question: What is the impact of envelope protein mutations on B-cell epitopes of Chikungunya virus (CHIKV) from the 2016 Alagoas outbreak in Brazil?

Findings: In this descriptive study, partial sequencing of the E1, E2, and 6K genes from 13 CHIKV-positive samples revealed 13 amino acid mutations. *In silico* analysis indicated that several mutations are located in surface-exposed regions, overlap known B-cell epitopes, and alter predicted conformational epitopes.

Meaning: These findings highlight the importance of genomic surveillance to inform vaccine, diagnostic, and therapeutic strategies by identifying mutations that may impact immune recognition.

✉To whom correspondence may be addressed. E-mail: enio.bassi@icbs.ufal.br; leticia.anderson@iqb.ufal.br

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

©2025 *Asian Pacific Journal of Tropical Medicine* Produced by Wolters Kluwer-Medknow.

How to cite this article: Taniele-Silva J, De Andrade Brandão J, De Oliveira MJTCC, De Souza SJM, Ferro JFG, De Lima MC, et al. Exploring the impact of envelope protein mutations on Chikungunya virus epitopes: Analysis of virus samples from the Alagoas State outbreak, Brazil. Asian Pac J Trop Med 2025; 18(6): 269-279.

Article history: Received 20 December 2024 Revision 7 June 2025
Accepted 20 June 2025 Available online 27 June 2025

first time on the American continent in 2013, leading to epidemic outbreaks in the Caribbean region and later spreading to countries in Central and South America[2]. The virus was first identified in Brazil in 2014 and has since emerged in the region[3]. In 2016, there was a Chikungunya outbreak, mainly in the northeastern region of Brazil, which accounted for 85% of the more than 270 000 probable reported cases[4]. The state of Alagoas had the highest incidence rate in the northeast region, with 548 cases per 100 000 inhabitants[4].

CHIKV is an enveloped virus with a single-stranded positive-sense RNA genome of approximately 11 800 nucleotides (nt)[5] that encodes nonstructural proteins (nsP1, nsP2, nsP3, and nsP4) and structural proteins (capsid, E3, E2, 6K, and E1)[6,7]. RNA viruses such as CHIKV have undergone rapid evolution[5,8]. This leads to a variety of genome mutations that contribute to heightened epidemic potential, the ability to be transmitted by new vectors[9], increased pathogenicity in animal models[10], resistance to antivirals[11] and evasion of the immune system[12,13]. Knowledge of genetic variation and virus mutations in Brazilian Northeast outbreak strains is currently limited[14-16]. Therefore, monitoring virus mutations is crucial.

Envelope proteins are the main targets of neutralizing antibodies[17,18]. Mutations in genes encoding important domains of these proteins can affect the interaction between antibodies and antigens because genetic variation in key antibody-binding regions may occur[19]. Therefore, mutations may reduce the sensitivity of immunodiagnostic tests, such as monoclonal antibody-based immunochromatographic assays[20-22].

This study aimed to investigate mutations in the CHIKV envelope genome region in samples from the 2016 outbreak in the State of Alagoas, Brazil. Additionally, the potential consequences of these mutations on B lymphocyte epitopes were examined through *in silico* analysis. These findings advance the understanding of the molecular epidemiology and interactions between CHIKV and the host humoral immune response, thus contributing to the design of immunodiagnostic assays, antibody therapies, and epitope-based vaccines.

2. Methods

2.1. Serum viral samples

The samples were provided by the Laboratório Central de Saúde Pública de Alagoas (LACEN/AL). This study was approved by the Universidade Federal de Alagoas Research Ethical Committee (C.A.A.E.) with the number 59229716.9.0000.5013 (approved on 6 October 2016), following National Council of Health of Brazil (CNS) resolution number 466/12. Viral RNA was isolated from 13 CHIKV-positive serum samples collected during the 2016 outbreak in the State of Alagoas, Brazil. The samples used in this study were obtained from patients with suspected CHIKV infection, presenting characteristic symptoms of the acute viremic phase of the disease, following patient management guidelines from the Brazilian Ministry of Health. Viral infection was confirmed by RT-qPCR using the specific primer set for the nonstructural protein NSP4, as previously described[23].

2.2. Viral RNA isolation, RT-PCR, and Sanger sequencing

Viral RNA was isolated from 200 μL of serum *via* the ReliaPrep™ Viral TNA Miniprep System (Promega, Madison, USA) according to the manufacturer's instructions. Reverse transcription was conducted with the ImProm-Ⅱ™ Reverse Transcription System (Promega, Madison, USA) using specific primers for the CHIKV envelope region, which were designed with PRIMER3 software[24] (Supplementary Table 1). The amplification of viral complementary DNA (cDNA) was performed by PCR using GoTaq® G2 Green Master Mix (Promega, Madison, USA) with specific primers for the E1, E2, and E3 genes (Figure 1A). The PCR cycling conditions included 35 cycles of denaturation at 94 ℃ for 30 s, annealing at 60 ℃ for 45 s, and extension at 72 ℃ for 1 min. The PCR products were analyzed by 1.5% agarose gel electrophoresis. The resulting bands were purified using a ReliaPrepTM DNA Clean-Up and Concentration System Kit (Promega, Madison, USA) and subsequently quantified with a Qubit dsDNA BR Assay Kit (Invitrogen). The purified PCR products were sequenced using Sanger's methodology in both directions with the BigDye Terminator Cycle Sequencing Ready Reaction on an ABI-Prism 3500 Genetic Analyzer (Applied Biosystems). All CHIKV partial genome sequences were assembled with BioEdit Sequence Alignment Editor version 7.2.5 to generate a contig for each analyzed segment and sample. The sequences were deposited in NCBI GenBank under accession numbers MN864707-MN864742.

2.3. In silico analysis of the nucleotide sequences

2.3.1. Phylogenetic analysis

Phylogenetic analysis was conducted using partial sequencing of the CHIKV *E1* gene (664 nt), which was subsequently aligned with 79 sequences retrieved from GenBank. Representative sequences from various CHIKV genotypes spanning different outbreaks, geographical locations, and years of isolation were included. Each selected sequence was designated with the GenBank accession number, locality, and year of isolation (Supplementary Table 2). Phylogenetic reconstruction was performed using the UPGMA statistical method and the Kimura-2 parameter model (K2) with 1 000 bootstrap replications. O'nyong-nyong virus (ONNV) (GenBank accession number AF079456) served as the outgroup.

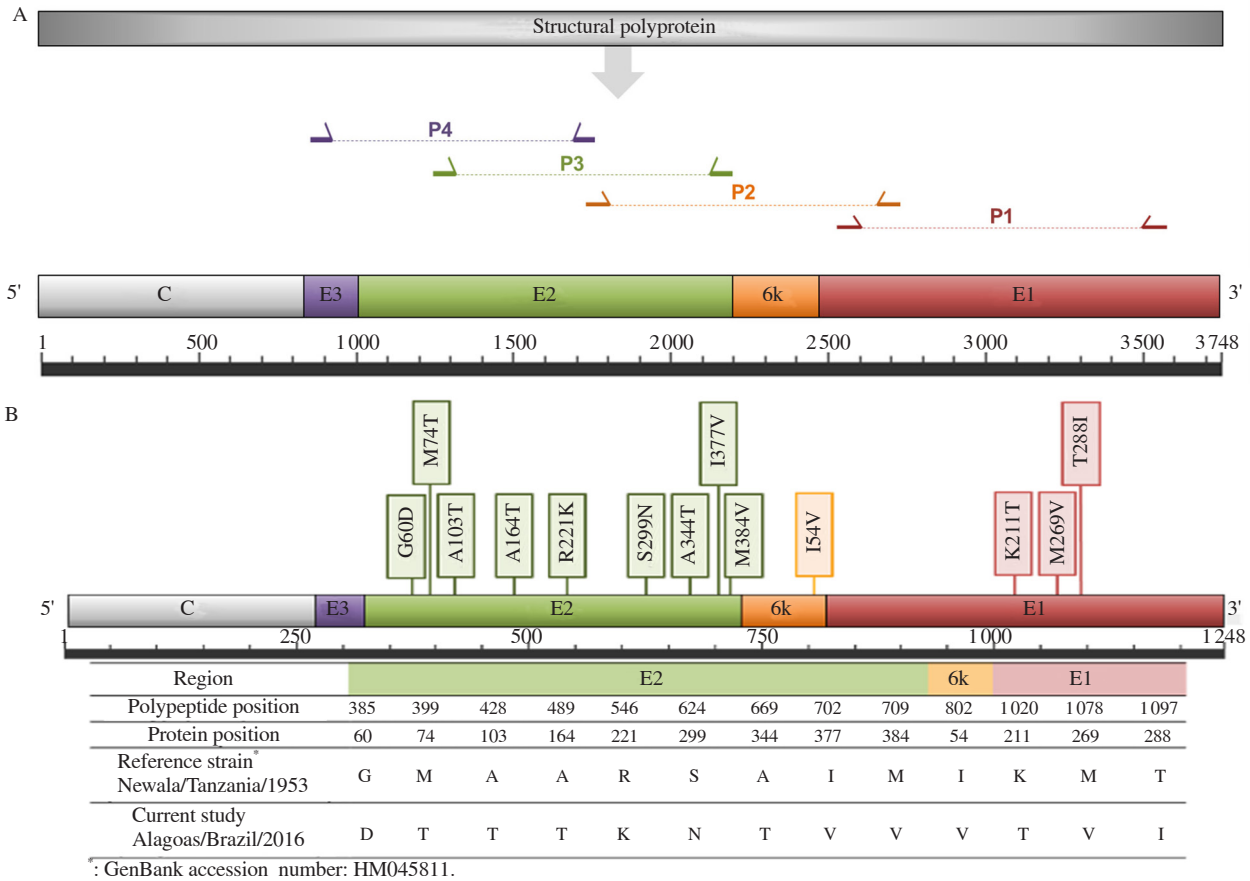
2.3.2. Mutation prospecting

The contig sequences of this study were aligned with a reference strain identified in Tanzania in 1953 (CHIKV strain Ross low-psg; GenBank: HM045811) using the online software MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/)[25]. Nucleotide mutations were identified using the online software InSites® (http://indra.mullins.microbiol.washington.edu/DIVEIN/insites.html) (Mullins Lab,

University of Washington, Seattle, Washington, USA, 2011). The alignments were subsequently translated *in silico* using MEGA software version 7.0 (MEGA7) (https://www.megasoftware.net/)[26].

2.3.3. Three–dimensional mapping of mutations

The amino acid sequences of the E1 and E2 proteins from the Alagoas/2016 and Tanzania/1953 strains (GenBank: HM01855) were used as targets for homology modeling *via* the SWISS-MODEL server[27]. SWISS-MODEL queries putative X-ray model proteins in the Protein Data Bank (PDB) to generate 3D models for all target sequences. The best homology models selected presented the highest coverage of the proteins with the amino acid sequences analyzed



Region	E2									6k	E1		
Polypeptide position	385	399	428	489	546	624	669	702	709	802	1020	1078	1097
Protein position	60	74	103	164	221	299	344	377	384	54	211	269	288
Reference strain* Newala/Tanzania/1953	G	M	A	A	R	S	A	I	M	I	K	M	T
Current study Alagoas/Brazil/2016	D	T	T	T	K	N	T	V	V	V	T	V	I

*: GenBank accession number: HM045811.

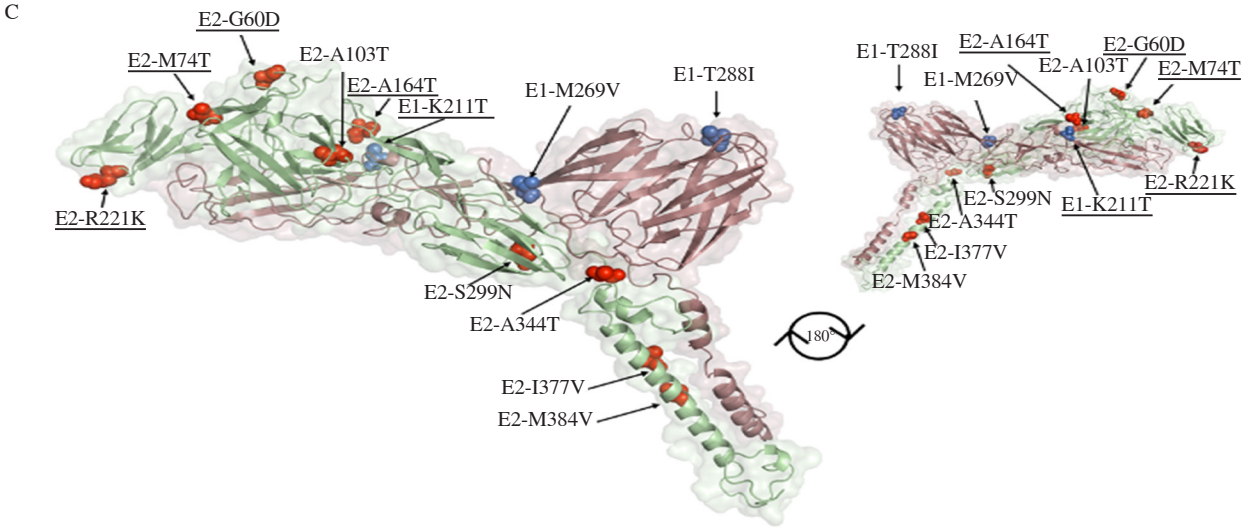


Figure 1. (A) Illustrative scheme of the primers annealing regions. The top graphite bar depicts the structural polyprotein. The half arrow pointing to the right indicates the sense primer, whereas the half arrow pointing to the left represents the antisense primer. The dotted lines indicate the expected product. The black bar illustrates the nucleotide positions. The colorful boxes correspond to the viral envelope proteins: E3 (purple), E2 (green), 6K (orange) and E1 (red). CHIKV structural glycoprotein localization of nonsynonymous mutations: (B) Amino acid mutations are positioned in the CHIKV polypeptide E2-6k-E1 and individually in structural proteins. The amino acids of the reference strain to be compared with those of the Alagoas strain (described in this study) are indicated. The mutations detected are boxed in the same color of the protein: E2 in green, 6K in orange, and E1 in red. (C) Structural mapping of mutations in the 3D structure of envelope glycoproteins. The PDB structure 6NK5 of E1 (brown) and E2 (green) was used to highlight the positions of the mutated residues detected. The blue dots correspond to mutated residues in glycoprotein E1, and the red dots correspond to mutated residues in glycoprotein E2. The residues that are in the exposed regions of the surface of the E1-E2 dimer are underlined.

in this study. Furthermore, the selection criteria considered the structures in the PDB with the highest resolution, highest QMEAN Disco Global value, and *Z* score values for the structural properties, ensuring the best fit to experimentally determined protein structure models and highest QMEAN values for the relevant residues. The models were obtained from templates with high resolution as follows: 3n41.1. C (3.01 Å) for E1, 3n41.1. B (3.01 Å) for E2. To identify solvent-exposed mutated residues within the models, nonsynonymous mutations were mapped onto three-dimensional structures using the PyMOL Molecular Graphics System, version 2.0[28].

To highlight the location of the mutated residues in the glycoprotein structure (Figure 1), we used 6NK5 electron cryomicroscopy of the Chikungunya virus structure, which delineates both the exposed and transmembrane domains in the virus-like particle.

2.3.4. Immunoinformatic analysis of mutations in the CHIKV sequence experimental epitope identification

To detect mutations that may be present in regions of experimentally validated epitopes, all epitopes described in the structural proteins of CHIKV available in the Immune Epitope Database (IEDB) (http://www.iedb.org/) were investigated[29]. The search targeted organisms identified as the CHIKV (ID: 37124) and the prototype of the S27 strain of the CHIKV (ID: 371094). The host, type of assay, and Major Histocompatibility Complex (MHC) were not specified in the search strategy. This analysis was conducted with epitopes deposited in the IEDB up to March 18, 2022.

2.3.5. Prediction of conformational B-cell epitopes

The prediction of conformational B-cell epitopes with mutations was verified using the online tool DiscoTope 2.0[30], which is available at the IEDB (http://tools.iedb.org/discotope/). Epitopes were predicted using modeled proteins from the SWISS-MODEL server and compared with ancestral strain protein models. Residues with a score higher than -3.7 (sensitivity: 0.47 and specificity: 0.75) were considered conformational epitopes[31].

2.3.6. Prediction of linear B-cell epitopes

The immunogenicity potential of linear epitopes was predicted using the reference strain sequence (Tanzania, 1953-GenBank:

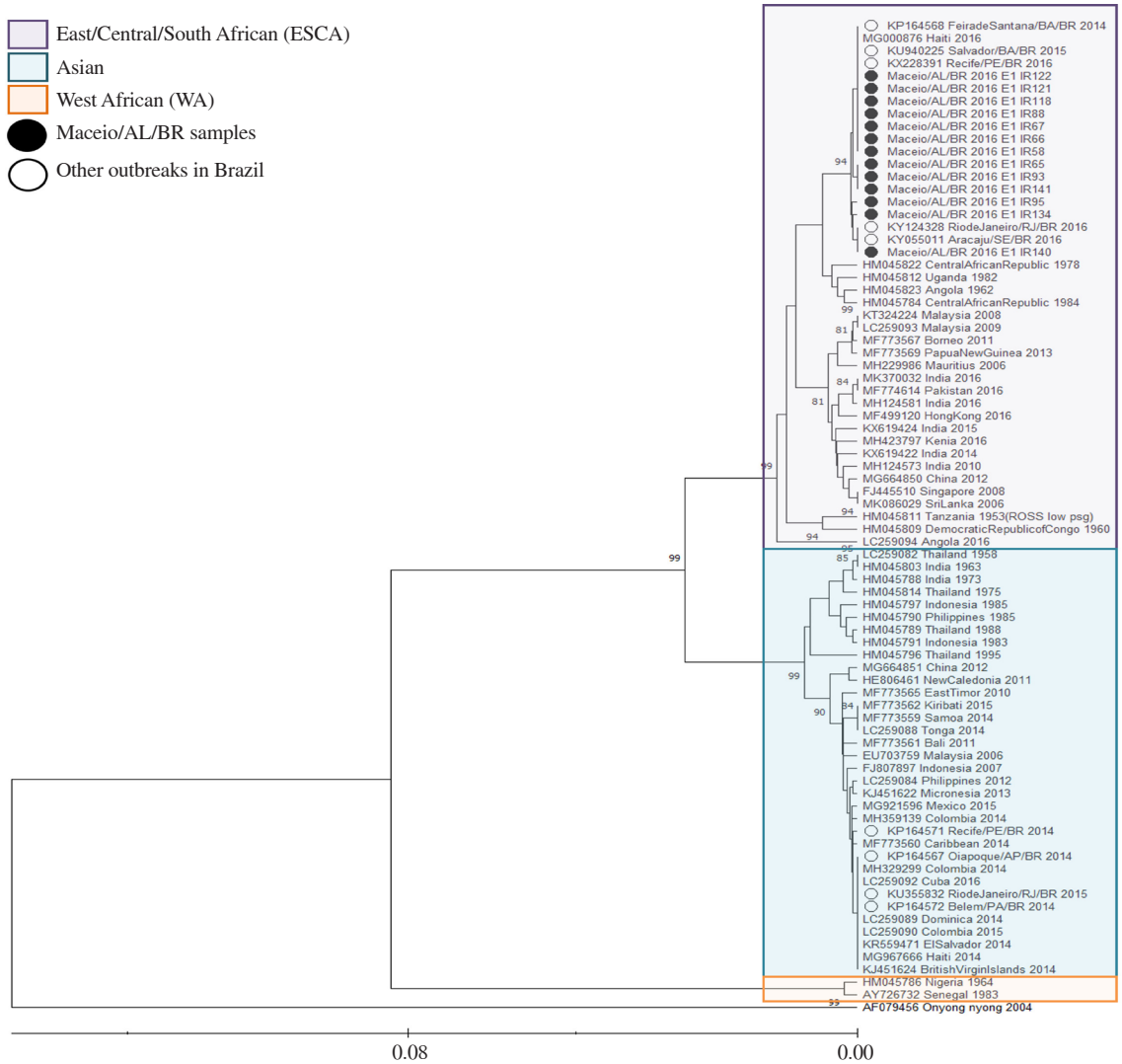


Figure 2. Phylogenetic tree based on CHIKV glycoprotein E1 partial sequence (664 nt). The UPGMA method in MEGA 7 software was utilized to infer the evolutionary history. Bootstrap values (>70%) for 1 000 replicates are displayed at the branch nodes. The evolutionary distances were calculated using the Kimura 2-parameter method and are expressed as the number of base substitutions per site. The 79 nucleotide sequences are labeled with the GenBank accession number, locality and year of the sample reported. The outgroup in the analysis was the O'nyong nyong virus sequence.

HM045811) and the mutations detected in this study. The prediction analysis was performed using different propensity scales available in the IEDB (http://tools.iedb.org/bcell/). These scales consider the propensity to be present in β loops (Chou & Fasman scale), flexibility (Karplus & Schulz scale), and hydrophilicity (Parker scale). Furthermore, the B-cell sequential epitope predictor algorithm (BepiPred 2.0) was used for epitope prediction. The tests were performed using the IEDB's standard parameters.

3. Results

3.1. ECSA CHIKV strain circulating in Alagoas/BR in 2016 and its amino acid mutations

The samples sequenced in this study were collected during the 2016 Chikungunya outbreak in the State of Alagoas, Brazil. Phylogenetic analysis revealed the presence of the ECSA genotype (Figure 2), which clustered with samples from other outbreaks in Brazil, such as Feira de Santana-Bahia (2014)[3], Salvador-Bahia (2015)[32], Recife-Pernambuco (2016) (GenBank code: KX228391.1), Rio de Janeiro-Rio de Janeiro (2016)[33], and Aracaju-Sergipe (2016) (GenBank code: KY055011)[34], indicating a common evolutionary origin.

Nucleotide sequence alignment with the first CHIKV strain (HM045811) isolated from Tanzania (1953) revealed 82 nucleotide substitutions in the coding sequences of envelope proteins E1, E2, and E3. Among these substitutions, 16 were nonsynonymous, leading to 13 amino acid mutations (Supplementary Files 1 and 2). The mutations detected in the proteins were E1 (K211T, M269V, and T288I), E2 (G60D, M74T, A103T, A164T, R221K, S299N, A344T, I377T, and M384V), and 6K (I54V) (Figure 1B).

Five mutations were found on the protein in the exposed regions of the viral particle (underlined residues in the text and Figure 1C), covering the E1 and E2 proteins. Specifically, at the E2 ectodomain, three mutations in the A domain (G60D, M74T, A103T), one mutation in the B domain (R221K), one mutation in the C domain (S299N), and one mutation in the acid-sensitive region (ASR) (A164T) were detected. The E2-A344T mutation occurs in a region that forms a stem before the transmembrane segment, which includes the E2-I377V and E2-M384V mutations located in an alpha helix region that interacts with the lipid membrane[35,36]. Mutations K211T and M269V occur in domain Ⅱ of the protein's E1 ectodomain, whereas T288I occurs in domain Ⅰ.

3.2. Mutations in experimentally validated discontinuous and linear epitopes

The search for experimental epitopes revealed 153 epitopes in the structural proteins of CHIKV, which were detected using different methods and studies. The nine mutations detected in this study were in seven B-linear epitopes, two T-linear epitopes, and nine discontinuous B-cell epitopes (Table 1). At positions E2-74, E2-103 and E2-384, the identified epitopes had residues different from those detected in the strains of this study (same amino acids as the Tanzania strain). Importantly, at positions E2-60, E2-164, E2-221, E2-299, E2-344, and 6K-54, the experimentally validated epitopes presented sequences identical to those of the strains in this study, revealing potential antigenic sites at the glycoprotein of the CHIKV Alagoas strain in B and T cells.

Table 1. Experimentally validated epitopes deposited in the IEDB database, whose sequences overlap positions of the mutations detected in this study.

Mutation	ID/Type	Epitope residues	Host/Positive assays	Ref.
E2-G60D	462266 B-cell discontinuous epitope	N7; V8; K10; A11; D59; D60; H62; E166; V169; M171; G194; T212; H232; K234; W235; R251; K252; G253; K254; I255	Host: *Homo sapiens* Positive assays: -Electron microscopy 3D structure -Biologic activity neutralization -Qualitative binding flow cytometry	[37]
	462515 B-cell discontinuous epitope	F6; N7; V8; Y9; K10; T12; T58; D59; D60; S61; H62; N193; G194; Q195; H232; K234	Host: *Homo sapiens* Positive assays: -Electron microscopy -3D structure	[18]
	462267 B-cell discontinuous epitope	N7; V8; K10; A11; R13; K57; T58; D59; D60; S61; D63; K66; H73; M74; P75; A76; D77; V169; M171; N193; G194; Q195; V229; N231; H232; K233; K234; W235; G253	Host: *Homo sapiens* Positive assays: -Electron microscopy 3D structure -Biological activity neutralization -Qualitative binding flow cytometry	[37]
	434425 B-cell discontinuous epitope	D385*; R393; G423; H495; M496; K558; K559	Host: *Homo sapiens* Positive assays: -Biological activity neutralization -Immuno staining qualitative binding	[38]
	434439 B-cell discontinuous epitope	T383; D384; D385*; R393; D396; I399; D402; T516; N518; K559	Host: *Homo sapiens* Positive assays: -Biological activity neutralization -ELISA qualitative binding -Immuno staining qualitative binding	[38]
	2134442 B-cell linear epitope	IQVSLQIGIKTDDSHDWT	Host: *Homo sapiens* Positive assays: -ELISA qualitative binding	[39]

Table 1 continues in next page.

Table 1. Continued.

Mutation	ID/Type	Epitope residues	Host/Positive assays	Ref.
E2-M74T	462267 B-cell discontinuous epitope	N7; V8; K10; A11; R13; K57; T58; D59; D60; S61; D63; K66; H73; M74; P75; A76; D77; V169; M171; N193; G194; Q195; V229; N231; H232; K233; K234; W235; G253	Host: *Homo sapiens* Positive assays: -Electronic microscopy 3D structure -Biological activity neutralization -Qualitative binding flow cytometry	[37]
	434421 B-cell discontinuous epitope	A11; D59; M74; N193; G194; T212; H232; W235	Host: *Mus musculus* Positive assays: -Biological activity neutralization -Immuno staining qualitative binding	[38]
	236399 B-cell discontinuous epitope	H398; I399*; H424; M496; Q520; T521; Y524; C526; G534; T537; N556; K558	Host: *Homo sapiens* Positive assays: -Binding assay dissociation constant KD -Binding assay on and off rate -Biological activity neutralization -Flow cytometry qualitative binding	[40]
E2-A103T	558231 B-cell linear epitope	CTITGTMGHFILARC + *AMID(C15)*	Host: *Homo sapiens* Positive assay: -Elisa qualitative binding	[19]
E2-A164T	226466 B-cell linear epitope	ATTEEIEVHMPPDTPDRT	Host: *Macaca fascicularis* Positive assay: -Elisa qualitative binding	[41]
E2-R221K	558343 B-cell linear epitope	VINNCKVDQCHAAVT + *AMID(T15)*	Host: *Homo sapiens* Positive assay: -Elisa qualitative binding	[19]
	1329042 B-cell discontinuous epitope	Q509, S510, G511, N512, V513, K514, D539, K540, V541, I542, N543, N544, K546	Host: *Mus musculus* -Bio-layer interferometry assay -Dissociation constant KD -Electron microscopy 3D structure -Biological activity neutralization -*In vivo* assay: pathogen burden after challenge after adoptive transfer	[42]
E2-S299N	237912 B-cell linear peptide	RNMGEEPNYQEEWVMHKK	Host: *Mus musculus* -Elisa qualitative binding	[43]
E2-A344T	993482 B-cell linear peptide	PTEGLEVTWGNNEPYKYWPQLSTNGT	Host: Mus musculus -ELISA qualitative binding	[44]
E2-M384V	993379 B-cell linear peptide	LLSMVGMAAGMCMCARRRCITPYELTPGATVPFL	Host: *Mus musculus* -ELISA qualitative binding	[44]
6K-I54V	737808 T-cell (CD8+) linear peptide	TLAFLAVMSV	Host: *Mus musculus* -IFNγ-secreting cell detection by intracellular cytokine staining -Cellular MHC/direct/fluorescence half maximal effective concentration (EC_{50})	[45]
	737939 T-cell (CD8+) linear peptide	VMSVGAHTV	Host: *Mus musculus* -IFNγ-secreting cell detection by intracellular cytokine staining -Cellular MHC/direct/fluorescence half maximal effective concentration (EC_{50})	[45]

The first column presents the mutation detected in this study. The following columns show the epitope ID, the epitope type, the epitope amino acids, the host, and positive assays, followed by the scientific literature. In the column "epitope residues", the amino acids sequenced at the Alagoas strain (2016) are highlighted in green, and the red ones are present at the wild strain. * - Position in the structural polyprotein. Data collection carried out until Dec 12th, 2022.

3.3. Discontinuous B-cell epitopes predicted with mutations

The analysis of conformational B-cell epitopes predicted possible surface antigen residues from ancestral and Alagoas strains. The number of predicted conformational B-cell epitopes varied with strain. Therefore, 9 conformational epitopes in the E1 glycoprotein and 155 epitopes in the E2 protein were predicted in the ancestral strain, whereas 7 epitopes in the E1 and 156 epitopes in E2 were predicted in the Alagoas strain, demonstrating that the mutations affected the number of predicted conformational epitopes. In the analysis, we detected conformational epitopes, with three epitopes observed in the E2 glycoprotein and one in the E1 glycoprotein of both the reference strain and the Alagoas strain (Table 2). The predicted epitopes may have a three-dimensional structure with or

without mutated residues, suggesting antigenic potential for both the reference and the Alagoas strains.

The experimentally validated epitope data (Table 1) support the previous demonstration of E2-G60D as a B-cell discontinuous epitope. These new findings indicate that E2-74 is a possible discontinuous epitope with variation to a threonine (Table 2), such as the experimental epitope presenting a methionine (Table 1). Additionally, E2-299 has been identified as part of a linear and potential conformational epitope. Furthermore, E1 at position 211, which could present a lysine or threonine (Table 2), was identified as a potential antigenic epitope, although it was not found among the experimental epitopes.

Table 2. Conformational epitopes identified at DiscoTope algorithm with reference strain and the Alagoas strain.

Protein	Position	Tanzânia-1953		Alagoas-2016	
		Residue	DiscoTope score	Residue	DiscoTope score
E2	60	GLY	2.386	ASP	2.683
	74	MET	-0.111	THR	-0.661
	299	SER	-0.607	ASN	0.613
E1	211	LYS	-2.979	THR	-3.541

The table shows the mutated protein, position, residue, and DiscoTope score (cutoff higher than -3.7). Residues out of cutoff were not represented in the table. Residues colored in blue are present in the Tanzania strain (1953) and residues colored in red are present in the sequence from Alagoas (2016).

3.4. Potential new linear B–cell epitopes affected by mutations

The prediction of B-cell epitopes on glycoproteins revealed linear sequences with antigenicity potential that have not yet been experimentally tested. Ten of the thirteen positions were predicted to belong to epitopes in at least one analysis (Table 3). Interestingly, the potential antigenicity of E2-74 was demonstrated as a sequential epitope considering the amino acid threonine from the Alagoas strains (Table 3), which was shown to be only part of a discontinuous epitope with the methionine from the reference strain (Table 1). Protein E1 was pointed with one linear epitope enclosing the methionine (Tanzania strain) at position 269 but not with the valine found in our study according to BepiPred 2.0 analysis. In addition, linear potential epitopes were identified enclosing residues 211 and 288 of the E1 glycoprotein with the reference and Alagoas strains (Table 3), which were not found experimentally in the IEDB database (Table 1). These findings indicate novel potential epitopes that could be explored in further studies. All the other mutations analyzed were found to be part of epitopes with at least one tool and with the amino acid of the reference strain.

Table 3. Predicted linear B cell epitopes analysis with the mutations detected in CHIKV Alagoas virus.

Protein	Position	Chou & Fasman		Karplus & Schulz		Parker		BepiPred 2.0	
		Tanzania (1953)	Alagoas (2016)	Tanzania (1953)	Alagoas (2016)	Tanzania (1953)	Alagoas (2016)	Tanzania (1953)	Alagoas (2016)
E2	60	^{57}KTDGSHD^{63}	^{57}KTDDSHD^{63}	^{57}KTDGSHD^{63}	^{57}KTDDSHD^{63}	^{57}KTDGSHD^{63}	^{57}KTDDSHD^{63}	^{56}IKTDGSHDW TKLRYMDNH MPADAER^{80}	^{56}IKTDDSHDW TKLRYMDNH TPADAER^{80}
	74	^{71}DNHMPAD^{77}	^{71}DNHTPAD^{77}	-	^{71}DNHTPAD^{77}	^{71}DNHMPAD^{77}	^{71}DNHTPAD^{77}	-	-
	103	-	-	-	-	-	-	-	-
	164	-	-	^{161}AATAEEI^{167}	^{161}AATTEEI^{167}	^{161}AATAEEI^{167}	^{161}AATTEEI^{167}	^{131}HDPPVIGREKF HSRPQHGKELPC STYVQSTAATAE EI^{167}	^{131}HDPPVIGR EKFHSRPQ HGKELPCSTYV QSTAATTE^{165}
	221	^{218}NNCRVDQ^{224}	^{218}NNCKVDQ^{224}	-	^{218}NNCKVDQ^{224}	^{218}NNCRVDQ^{224}	^{218}NNCKVDQ^{224}	-	-
	299	^{296}SYRSMGE^{302}	^{296}SYRNMGE^{302}	^{296}SYRSMGE^{302}	^{296}SYRNMGE^{302}	^{296}SYRSMGE^{302}	^{296}SYRNMGE^{302}	-	^{299}NMGEEPNY QEEWVTHK^{314}
	344	^{341}QLSANGT^{347}	^{341}QLSTNGT^{347}	^{341}QLSANGT^{347}	^{341}QLSTNGT^{347}	^{341}QLSANGT^{347}	^{341}QLSTNGT^{347}	^{330}WGNNEPYKYW PQLSANGTAHG^{350}	^{330}WGNNEPYK YWPQLSTNGT AHG^{350}
	377	-	-	-	-	-	-	-	-
	384	-	-	-	-	-	-	-	-
6K	54	^{51}VMSIGAH^{57}	^{51}VMSVGAH^{57}	-	-	^{51}VMSIGAH^{57}	^{51}VMSVGAH^{57}	-	-
E1	211	^{208}PESKDVY^{214}	^{208}PESTDVY^{214}	^{208}PESKDVY^{214}	^{208}PESTDVY^{214}	^{208}PESKDVY^{214}	^{208}PESTDVY^{214}	^{183}DVYNMDYPPFG AGRPGQFGDIQSR TPESKDVYAN^{216}	^{183}DVYNMDYPP FGAGRPGQFG DIQSRTPESTD VYAN^{216}
	269	-	-	-	-	-	-	^{266}VRAMNC^{271}	-
	288	-	-	-	-	-	-	^{285}AAFTRVVDA PSL^{296}	^{285}AAFIRVVD APSL^{296}

The protein, the mutation position and the results of the propensity scales and tested algorithm are available for the reference lineage and the lineage from Alagoas, with the start and end position of the epitope superscript. Residues colored in blue are present in the Tanzania strain (1953) and residues colored in red are present in the sequence from Alagoas (2016). The tests were performed using the standard parameters of the IEDB.

4. Discussion

This study describes the molecular characterization of CHIKV envelope proteins during the 2016 outbreak in the State of Alagoas, Brazil, and their relevance to immunological epitopes. The CHIKV analyzed in this work belonged to the ECSA strain in all the samples, as shown in previous studies[14,15]. In addition to identifying mutations in the circulating strain, we reported nonsynonymous substitutions in the envelope glycoproteins E2 and E1 compared with those of an ancestral strain and, for the first time, demonstrated these mutations in the circulating CHIKV genome during the 2016 outbreak in Alagoas and their relevance to immune recognition. To date, studies have shown that CHIKV mutations contribute to viral dissemination by enhancing fitness in the vector, emphasizing the necessity of genomic monitoring. The clinical manifestations associated with the viral mutations have been linked to infection load in animal models. On the other hand, studies focusing on virus epitopes and how they can affect the mammalian host recognition are still being explored, mainly in the search for potential vaccines and antiviral treatments.

In this work, thirteen amino acid mutations in the CHIKV envelope proteins were pointed. The molecular evolution of CHIKV occurs gradually, with an estimated rate of 2.78×10^{-4} substitutions per site per year, suggesting that substantial genetic variations require prolonged periods of virus circulation[46]. Consequently, samples obtained from the same location within a short timeframe have a low probability of exhibiting significant evolutionary divergence.

The E1-K211T mutation detected in this study was first reported during an outbreak in Rio de Janeiro in 2016[33]. A mutation in the same position, E1-K211N, was detected during an outbreak in southwest India during the 2006 outbreak and in Sri Lanka in 2008, a locality in which the main Chikungunya vector is *Aedes (Ae.) aegypti*[47,48]. Another different mutation, E1-K211E, was detected in Indian strains of the ECSA genotype from Delhi, Andhra Pradesh, and Tamil Nadu[49–51] and in isolates of travelers in France returning from India endemic areas[52], often reported with E2-V264A[53]. E1-K211E is associated with a 6-fold increase in the viral titer in the *Ae. aegypti* mosquito intestine and a 12-fold increase in saliva, suggesting that it modulates viral infectivity and transmission[53]. When combined with E2-V264A, as reported in several genetic studies, infection, spread, and transmission are even greater (13-fold, 15-fold, and 62-fold, respectively), indicating an epistatic interaction between both mutations[53].

The E1-E211K mutation has been identified as a key amino acid within linear epitopes that impacts the effectiveness of cross-neutralization among distinct CHIKV genotypes[19]. In our study, we detected a threonine substitution at this position (E1-K211T), which may also affect the recognition of neutralizing antibodies. This mutation has been associated with increased swelling at the infection site, as well as higher muscle and serum viral titers, leading to enhanced viral pathogenicity in a mouse model[54]. However, its association with clinical manifestations in humans has not yet been demonstrated.

The E2 protein contains immunodominant epitopes recognized by neutralizing antibodies[55], and plays a key role in cell receptor recognition[56]. Most of the detected mutations (nine in total) are functionally relevant to viral infection, as they occur in domains and residues of the E2-E1 heterodimers that interact with the Mxra8 receptor-an entry receptor for arthritogenic alphaviruses, including CHIKV[56]. All residues previously described as binding sites to MXRA8 were identified in our study as mutation sites, including E2-74, E2-221, and E1-211, all of which are located in the exposed virus ectodomain. These mutations may influence the virus entry mechanism and should be considered in the development of entry inhibitors.

Specifically, the E2-M74T mutation is in the wing region of domain A[36]. The wild-type residue E2-M74 was detected within the conformational epitope targeted by the murine antibody CHK-152[38], whose mechanism involves blocking viral fusion, exerting neutralizing action against all three CHIKV genotypes *in vitro*, and protecting mice from lethal infection by reducing the viral load after passive transfer[57]. Residue E2-M74 is also part of the target epitope of antibodies 1H12 and 5M16, which also bind to residue E2-G60D[37,38], indicating its importance in the mechanisms of binding and neutralizing CHIKV.

The detected E2-60D is within linear and conformational epitopes to B cells in activity neutralization assays, immunostaining qualitative binding assays, and ELISA, which shows its antigenic importance. The E2-G60D mutation is located in a region termed "the Wings" at the top of the A domain of the E2 glycoprotein[36]. Domain A is known to be more structurally stable and exposed in the mature viral conformation, favoring the cell entry mechanism. E2-G60D increases CHIKV infectivity in the intestinal cells of *Ae. aegypti* mosquitoes and acts synergistically with E2-I211T and E1-A226V, increasing CHIKV infectivity in *Ae. albopictus* mosquitoes[58]. The literature demonstrates that this residue is part of a conformational epitope that strongly binds to neutralizing antibodies (3N23 and 1H12), which can completely neutralize CHIKV strains from different genotypes (IOL, WA, and Asian) when administered prophylactically in mice[38]. The neutralizing antibody MAb 8B10, which inhibits the binding of the virus to the cell receptor, also binds to a conformational epitope containing the E2-60D residue[18]. The E2-60D residue is part of the target epitopes for mAbs 4J21 and 5M16[37], inducing therapeutic activity in immunodeficient mice and inhibiting viral fusion with target cell membranes[38].

The E2-A103T mutation was previously detected in isolates from patients in Rio de Janeiro, Brazil, during the 2016 outbreak and is present in all Brazilian isolates[33]. The wild-type amino acid E2-A103 was detected as part of a B-cell epitope targeted to human antibodies and is suggested as a potential candidate for serological assays because it can bind to the sera of patients with ECSA and Asian CHIKV genotypes or as a vaccine candidate owing to its strong neutralizing action[19]. Thus, the presence of a mutation in this position must be considered because the substitution may prevent the antigenic properties of that region.

The E2-A164T mutation was previously reported in the molecular signature of the IOL sublineage in Indian isolates and was detected in 2000 in mosquito isolates in Yawat and human sample isolates during the 2006 Indian outbreak[48]. This mutation was also detected in 2010 isolates from New Delhi, along with mutations E1-M269V, E2-S299N, and E2-A344T[50], which were also detected in this study.

The E2-I377V and M384V mutations found in our study were also detected in Indian isolates from Kolkata (1963) and Barsi (1973) (Asian genotype), but there is no clear evidence of recombination between the ECSA and Asian genotypes[48].

The immunization of mice with different predicted and synthesized peptides covering residues E2-60D, E2-M74, E2-S299, E2-A164, E2-221K, and E2-A344 induced immunoreactivity, high antibody titers, and memory B cells after challenge with the E2 protein, demonstrating that epitope predictions are useful for identifying potential immunogenic peptides[59]. However, it is noteworthy to note that some of the peptides were synthesized with wild-type residues, and additional experimental studies are needed to determine whether these mutations alter the immune recognition.

Our findings highlight the importance of monitoring mutations in CHIKV structural proteins, as they may influence immune recognition and diagnostic accuracy. While our study identified potential epitope-altering mutations, further validation through experimental approaches-such as monoclonal antibody binding assays, neutralization studies, and structural analyses-is essential to determine their functional impact. Additionally, continuous genomic surveillance is crucial to assess whether these mutations affect the sensitivity and specificity of serological and molecular diagnostic tests, ensuring their reliability across different viral variants.

Taken together, our results underscore the need for a comprehensive understanding of CHIKV genetics for the development of vaccines and antiviral therapies. The design of peptide-based vaccines remains a promising approach for controlling Chikungunya outbreaks. However, this efficacy depends directly on selecting peptides that encompass immunogenic epitopes representative of circulating viral strains, rather than relying on a single strain, due to the considerable genetic variation of this virus.

Conflict of interest statement

The authors declare that they have no conflicts of interest.

Acknowledgments

We would like to thank all the technical support of the Laboratorio Central de Saude Publica do Estado de Alagoas (LACEN/AL), Maceio, Alagoas, Brazil, and Ana Rachel V. Lima from Laboratorio de Pesquisas em Virologia e Imunologia (LAPEVI).

Funding

This study was supported by Decit/SCTIE-Ministério da Saúde, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Alagoas (FAPEAL) and Secretaria de Estado da Saúde de Alagoas (SESAU-AL) [PPSUS 60030 000841/2016].

Authors' contributions

Concepts, design, experimental studies, data analysis and manuscript preparation: JTS. Experimental studies: MJTCCO, SJMS, JFGF and MCL. Experimental studies and manuscript editing: JAB. Manuscript review: ASJ. Concepts, design, experimental studies, data analysis and manuscript review: EJB and LA.

References

[1] Ketkar H, Herman D, Wang P. Genetic determinants of the re-emergence of arboviral diseases. *Viruses* 2019; **11**(2): 150.

[2] Leparc-Goffart I, Nougairede A, Cassadou S, Prat C, de Lamballerie X. Chikungunya in the Americas. *Lancet* 2014; **383**(9916): 514.

[3] Nunes MRT, Faria NR, de Vasconcelos JM, Golding N, Kraemer MU, de Oliveira LF, et al. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med* 2015; **13**(1): 102.

[4] Ministério da Saúde do Brasil. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 50, 2017. *Boletim Epidemiológico* 2018; **48**(45): 1-13.

[5] Khan AH, Morita K, del Carmen Parquet M, Hasebe F, Mathenge EGM, Igarashi A. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. *J Gen Virol* 2002; **83**(12): 3075-3084.

[6] Simizu B, Yamamoto K, Hashimoto K, Ogata T. Structural proteins of Chikungunya virus. *J Virol* 1984; **51**(1): 254-258.

[7] Powers AM, Brault AC, Shirako Y, Strauss EG, Kang W, Strauss JH, et al. Evolutionary relationships and systematics of the alphaviruses. *J Virol* 2001; **75**(21): 10118-10131.

[8] Holmes EC. The comparative genomics of viral emergence. *PNAS* 2010; **107**(1): 1742-1746.

[9] Tsetsarkin KA, Vanlandingham DL, Mcgee CE, Higgs S. A single mutation in Chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog* 2007; **3**(12): 1895-1906.

[10]Hawman DW, Carpentier KS, Fox JM, May NA, Sanders W, Montgomery SA, et al. Mutations in the E2 glycoprotein and the 3' untranslated region enhance Chikungunya virus virulence in mice. *J Virol* 2017; **91**(20): 1-17.

[11]Delang L, Guerrero NS, Tas A, Querat G, Pastorino B, Froeyen M, et al. Mutations in the chikungunya virus non-structural proteins cause resistance to favipiravir (T-705), a broad-spectrum antiviral. *J Antimicrob Chemother* 2014; **69**(10): 2770-2784.

[12]Kam YW, Lee WWL, Simarmata D, Harjanto S, Teng TS, Tolou H, et al. Longitudinal analysis of the human antibody response to Chikungunya virus infection: Implications for serodiagnosis and vaccine development. *J Virol* 2012; **86**(23): 13005-13015.

[13]Priya R, Dhanwani R, Patro IK, Rao PVL, Parida MM. Differential regulation of TLR mediated innate immune response of mouse neuronal cells following infection with novel ECSA genotype of Chikungunya virus with and without E1: A226V mutation. *Infect Genet Evol* 2013; **20**: 396-406.

[14]de Lira Tanabe EL, Tanabe ISB, Dos Santos EC, da Silva Marques JP, Borges AA, de Lima MC, et al. Report of East-Central South African Chikungunya virus genotype during the 2016 outbreak in the Alagoas State, Brazil. *Rev Inst Med Trop Sao Paulo* 2018; **60**: e19. doi: 10.1590/s1678-9946201860019.

[15]Charlys da Costa A, Thézé J, Cavalcante Vasconcelos Komninakis S, Lopes Sanz-Duro R, Lemos Felinto MR, Corrêa Moura LC, et al. Spread of Chikungunya virus East/Central/South African genotype in Northeast Brazil. *Emerg Infect Dis* 2017; **23**(10): 1742-1744.

[16]Dos Passos Cunha M, Dos Santos CA, de Lima Neto DF, Schanoski AS, Pour SZ, Passos SD, et al. Outbreak of chikungunya virus in a vulnerable population of Sergipe, Brazil—A molecular and serological survey. *J Clin Virol* 2017; **97**: 44-49.

[17]Kam YW, Simarmata D, Chow A, Her Z, Teng TS, Ong EKS, et al. Early appearance of neutralizing immunoglobulin G3 antibodies is associated with Chikungunya virus clearance and long-term clinical protection. *J Infect Dis* 2012; **205**(7): 1147-1154.

[18]Porta J, Prasad M, Wang C, Akahata W, Ng LFP, Rossmann G. Structural studies of Chikungunya virus-like particles complexed with human antibodies: Neutralization and cell-to-cell transmission. *J Virol* 2016; **90**(3): 1169-1177.

[19]Chua C, Sam I, Merits A, Chan Y. Antigenic variation of East/Central/South African and Asian Chikungunya virus genotypes in neutralization by immune sera. *PLoS Negl Trop Dis* 2016; **10**(8): e0004960. doi: 10.1371/journal.pntd.0004960.

[20]Tuekprakhon A, Nakayama EE, Bartholomeeusen K, Puiprom O, Sasaki T, Huits R, et al. Variation at position 350 in the Chikungunya virus 6K-E1 protein determines the sensitivity of detection in a rapid E1-antigen test. *Sci Rep* 2018; **8**(1) :1094.

[21]Okabayashi T, Sasaki T, Masrinoul P, Chantawat N, Yoksan S, Nitatpattana N, et al. Detection of Chikungunya virus antigen by a novel rapid immunochromatographic test. *J Clin Microbiol* 2015; **53**(2): 382-388.

[22]Yap G, Pok KY, Lai YL, Hapuarachchi HC, Chow A, Leo YS, et al. Evaluation of Chikungunya diagnostic assays: Differences in sensitivity of serology assays in two independent outbreaks. *PLoS Negl Trop Dis* 2010; **4**(7): e753.

[23]Gregianini TS, Ranieri T, Favreto C, Nunes ZMA, Tumioto Giannini GL, Sanberg ND, et al. Emerging arboviruses in Rio Grande do Sul, Brazil: Chikungunya and Zika outbreaks, 2014-2016. *Rev Med Virol* 2017; **27**(6). doi: 10.1002/rmv.1943.

[24]Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res* 2012; **40**(15): e115-e115.

[25]Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**(5): 1792-1797.

[26]Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets brief communication. *Mol Biol Evol* 2016; **33**(7): 1870-1874.

[27]Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018; **46**(W1): W296-W303. doi: 10.1093/nar/gky427.

[28]Schrödinger LLC, DeLano WL. *The PyMOL molecular graphics system, Version 2.0. 2015.* [Online]. Available from: https://www.pymol.org/. [Accessed on 20 June 2023].

[29]Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* 2019; **47**: 339-343.

[30]Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006; **15**(11): 2558-2567.

[31]Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Comput Biol* 2012; **8**(12): e1002829. doi: 10.1371/journal.pcbi.1002829.

[32]Lyra P, Campos G, Bandeira I, Sardi S, Costa L, Santos F, et al. Congenital Chikungunya virus infection after an outbreak in Salvador, Bahia, Brazil. *AJP Rep* 2016; **6**(3): e299-e300.

[33]Cunha MS, Schnellrath LC, Luiza M, Medaglia G, Casotto ME, Albano RM, et al. Autochthonous transmission of East/Central/South African

genotype Chikungunya virus, Brazil. *Emerg Infect Dis* 2017; **23**(10): 1737-1739.

[34]Costa-da-Silva AL, Ioshino RS, Petersen V, Lima AF, Cunha M dos P, Wiley MR, et al. First report of naturally infected *Aedes aegypti* with chikungunya virus genotype ECSA in the Americas. *PLoS Negl Trop Dis* 2017; **11**(6): e0005630.

[35]Mukhopadhyay S, Zhang W, Gabler S, Chipman PR, Strauss EG, Strauss JH, et al. Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure* 2006; **14**(1): 63-73.

[36]Voss JE, Vaney MC, Duquerroy S, Vonrhein C, Girard-Blanc C, Crublet E, et al. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nat Lett* 2010; **468**(7324): 709-712.

[37]Long F, Fong RH, Austin SK, Chen Z, Klose T, Fokine A, et al. Cryo-EM structures elucidate neutralizing mechanisms of anti-chikungunya human monoclonal antibodies with therapeutic activity. *PNAS* 2015; **112**(45): 13898-13903.

[38]Smith SA, Silva LA, Fox JM, Diamond MS, Dermody TS, Crowe JE, et al. Isolation and characterization of broad and ultrapotent human monoclonal antibodies with therapeutic activity against Chikungunya virus. *Cell Host Microbe* 2015; **18**(1): 86-95.

[39]Silva JPDC, Cunha MDP, Pour SZ, Hering VR, Neto DFL, Zanotto PMA. Chikungunya virus E2 structural protein B-cell epitopes analysis. *Viruses* 2022; **14**(8): 1839.

[40]Fong RH, Banik SSR, Mattia K, Barnes T, Tucker D, Liss N, et al. Exposure of epitope residues on the outer face of the chikungunya virus envelope trimer determines antibody neutralizing efficacy. *J Virol* 2014; **88**(24): 14364-14379.

[41]Kam YW, Lee WWL, Simarmata D, Le Grand R, Tolou H, Merits A, et al. Unique epitopes recognized by antibodies induced in Chikungunya virus-infected non-human primates: Implications for the study of immunopathology and vaccine development. *PLoS One* 2014; **9**(4): e95647.

[42]Zhou QF, Fox JM, Earnest JT, Ng TS, Kim AS, Fibriansah G, et al. Structural basis of Chikungunya virus inhibition by monoclonal antibodies. *Proc Natl Acad Sci U S A* 2020; **117**(44): 27637-27645.

[43]Her Z, Teng T, Tan JJ, Teo T, Kam Y, Lum F, et al. Loss of TLR3 aggravates CHIKV replication and pathology due to an altered virus-specific neutralizing antibody response. *EMBO Mol Med* 2015; **7**(1): 24-41.

[44]Basu R, Zhai L, Rosso B, Tumban E. Bacteriophage Qβ virus-like particles displaying Chikungunya virus B-cell epitopes elicit high-titer E2 protein antibodies but fail to neutralize a Thailand strain of Chikungunya virus. *Vaccine* 2020; **38**(11): 2542-2550.

[45]Lorente E, Barriga A, García-Arriaza J, Lemonnier FA, Esteban M, López D. Complex antigen presentation pathway for an HLA-A*0201-restricted epitope from Chikungunya 6K protein. *PLoS Negl Trop Dis* 2017; **11**(10): e0006036. doi: 10.1371/journal.pntd.0006036.

[46]Deeba F, Haider MSH, Ahmed A, Tazeen A, Faizan MI, Salam N, et al. Global transmission and evolutionary dynamics of the Chikungunya virus. *Epidemiol Infect* 2020; **148**: e63.

[47]Hapuarachchi HC, Bandara KBAT, Sumanadasa SDM, Hapugoda MD, Lai Y, Lee K, et al. Re-emergence of Chikungunya virus in South-east Asia: Virological evidence from Sri Lanka and Singapore. *J GenVirol* 2010; **91**(Pt 4): 1067-1076. doi: 10.1099/vir.0.015743-0.

[48]Arankalle VA, Shrivastava S, Cherian S, Gunjikar RS, Walimbe AM, Jadhav SM, et al. Genetic divergence of Chikungunya viruses in India (1963-2006) with special reference to the 2005-2006 explosive epidemic. *J Gen Virol* 2007; **88**(Pt 7): 1967-1976. doi: 10.1099/vir.0.82714-0.

[49]Kumar CVMN, Gopal DVRS. Reemergence of Chikungunya virus in Indian Subcontinent. *Indian J Virol* 2010; **21**(1): 8-17.

[50]Shrinet J, Jain S, Sharma A, Singh SS, Mathur K, Rana V, et al. Genetic characterization of Chikungunya virus from New Delhi reveal emergence of a new molecular signature in Indian isolates. *Virol J* 2012; **9**: 100. doi: 10.1186/1743-422X-9-100.

[51]Sumathy K, Ella KM. Genetic diversity of Chikungunya virus, India 2006-2010: Evolutionary dynamics and serotype analyses. *J Med Virol* 2012; **84**: 462-470.

[52]Grandadam M, Caro V, Plumet S, Thiberge J, Souarès Y, Failloux A, et al. Chikungunya virus, Southeastern France. *Emerg Infect Dis* 2011; **17**(5): 910-913.

[53]Agarwal A, Kumar A, Sukumaran D, Parida M, Kumar P. Two novel epistatic mutations (E1:K211E and E2:V264A) in structural proteins of Chikungunya virus enhance fitness in *Aedes aegypti*. *Virology* 2016; **497**: 59-68.

[54]Rangel MV, McAllister N, Dancel-Manning K, Noval MG, Silva LA, Stapleford KA. Emerging Chikungunya virus variants at the E1-E1 interglycoprotein spike interface impact virus attachment and inflammation. *J Virol* 2022; **96**(4): 1586-1607.

[55]Kam Y, Lum F, Teo T, Lee WWL, Simarmata D, Harjanto S, et al. Early neutralizing IgG response to Chikungunya virus in infected patients targets a dominant linear epitope on the E2 glycoprotein. *EMBO Mol Med* 2012; **4**(4): 330-343.

[56]Basore K, Kim AS, Nelson CA, Zhang R, Smith BK, Uranga C, et al. Cryo-EM structure of Chikungunya virus in complex with the Mxra8 receptor. *Cell* 2019; **177**(7): 1725-1737.e16.

[57]Pal P, Dowd KA, Brien JD, Edeling MA, Gorlatov S, Johnson S, et al. Development of a highly protective combination monoclonal antibody therapy against Chikungunya virus. *PLoS Pathog* 2013; **9**(4): e1003312. doi: 10.1371/journal.ppat.1003312.

[58]Tsetsarkin KA, Mcgee CE, Volk SM, Vanlandingham DL, Weaver SC, Higgs S. Epistatic roles of E2 glycoprotein mutations in adaption of Chikungunya virus to *Aedes albopictus* and *Ae. Aegypti* mosquitoes. *PLoS One* 2009; **4**(8): e6835. doi: 10.1371/journal.pone.0006835.

[59]Rao D, KumarNagar P, Verma P, Joshi G, Singh A. Mapping and immunological response of immunodominant B and T cell epitopes of E2 glycoprotein of chikungunya virus. *MOJ Immunol* 2016; **4**(1): 1-8.

Publisher's note

The Publisher of the *Journal* remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Edited by Zhang Q, Lei Y, Pan Y

>lcl|MN864739.1_cds_QQL93927.1_1 [protein=E3-E2 envelope protein] [protein_id=QQL93927.1] [location=<1..>438] [gbkey=CDS]

---TAC
TTAGCTCACTGTCCCGACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCATTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGA
ATAAAGACGGATGATAGCCACGATTGGACCAAGCTGCGTTACATGGACAATCACACGCCA
GCGGACGCAGAGAGGGCGGGGCTATTTGTAAGAACATCAGCACCGTGCACGATTACTGGA
ACAATGGGACACTTCATCCTGACCCGATGTCCGAAAGGGGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATCAGTCACTCATGTACGCACCCATTTCACCATGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAAAGCACC---

>lcl|MN864740.1_cds_QQL93928.1_1 [protein=E3-E2 envelope protein]
[protein_id=QQL93928.1] [location=<1..>672] [gbkey=CDS]

----------GAG
GACAACGTCATGAGACCTGGGTACTATCAGCTGCTACAAGCATCCTTAACATGTTCTCCC
CACCGTCAGCGACGCAGTACCAAGGACAACTTCAATGTCTATAAAGCCACAAGACCGTAC
TTAGCTCACTGTCCCGACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCATTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGA
ATAAAGACGGATGATAGCCACGATTGGACCAAGCTGCGTTACATGGACAACCACACGCCA
GCGGACGCAGAGAGGGCGGGGCTATTTGTAAGAACATCAGCACCGTGCACGATTACTGGA
ACAATGGGACACTTCATCCTGACCCGATGTCCGAAAGGGGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATCAGTCACTCATGTACGCACCCATTTCACCATGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGGACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAA
ACGGTGCGG----------


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864741.1_cds_QQL93929.1_1 [protein=E3-E2 envelope protein]
[protein_id=QQL93929.1] [location=<1..>729] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------CTACGCATGCTTGAG
GACAACGTCATGAGACTTGGGTACTATCAGCTGCTACAAGCATCCTTAACATGTTCTCCC
CACCGTCAGCGACGCAGTACCAAGGACAACTTCAATGTCTATAAAGCCACAAGACCGTAC
TTAGCTCACTGTCCCGACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCATTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGA
ATAAAGACGGATGATAGCCACGATTGGACCAAGCTGCGTTACATGGACAACCACACGCCA
GCGGACGCAGAGAGGGCGGGGCTATTTGTAAGAACATCAGCACCGTGCACGATTACTGGA
ACAATGGGACACTTCATCCTGACCCGATGTCCGAAAGGGGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATCAGTCACTCATGTACGCACCCATTTCACCACGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCAAATGAAGGACTAATAACTACA------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864736.1_cds_QQL93924.1_1 [protein=E3-E2 envelope protein]
[protein_id=QQL93924.1] [location=<1..>741] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------TGCTGCTATGAAAGGGAACCAGAGGAAACCCTACGCATGCTTGAG
GACAACGTCATGAGACCTGGGTACTATCAGCTGCTACAAGCATCCTTAACATGTTCTCCC
CACCGTCAGCGACGCAGTACCAAGGACAACTTCAATGTCTATAAAGCCACAAGACCGTAC
TTAGCTCACTGTCCCGACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCATTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGA
ATAAAGACGGATGATAGCCACGATTGGACCAAGCTGCGTTACATGGACAACCACACGCCA
GCGGACGCAGAGAGGGCGGGGCTATTTGTAAGAACATCAGCACCGTGCACGATTACTGGA
ACAATGGGACACTTCATCCTGACCCGATGTCCGAAAGGGGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATCAGTCACTCATGTACGCACCCATTTCACCATGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCAAAT----------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864742.1_cds_QQL93930.1_1 [protein=E3-E2 envelope protein]
[protein_id=QQL93930.1] [location=<1..>534] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------CATAGTCCCGTAGCATTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGA
ATAAAGACGGATGATAGCCACGATTGGACCAAGCTGCGTTACATGGACAACCACACGCCA
GCGGACGCAGAGAGGGCGGGGCTATTTGTAAGAACATCAGCACCGTGCACGATTACTGGA
ACAATGGGACACTTCATCCTGACCCGATGTCCGAAAGGGGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATCAGTCACTCATGTACGCACCCATTTCACCATGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCA---------------------------
------------------------------------------------------------
```


>lcl|MN864737.1_cds_QQL93925.1_1 [protein=E3-E2 envelope protein] [protein_id=QQL93925.1] [location=<1..>708] [gbkey=CDS]

--ACATGTTCTCCC
CACCGTCAGCGACGCAGTACCAAGGACAACTTCAATGTCTATAAAGCCACAAGACCGTAC
TTAGCTCACTGTCCCGACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCATTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATCGGA
ATAAAGACGGATGATAGCCACGATTGGACCAAGCTGCGTTACATGGACAACCACACGCCA
GCGGACGCAGAGAGGGCGGGGCTATTTGTAAGAACATCAGCACCGTGCACGATTACTGGA

ACAATGGGACACTTCATCCTGACCCGATGTCCGAAAGGGGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATCAGTCACTCATGTACGCACCCATTTCACCATGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCAAATGAAGGACTAATAACTACAGACAAA
GTGATTAATAACTGCAAAGTTGATCAATGTCATGCC------------------------

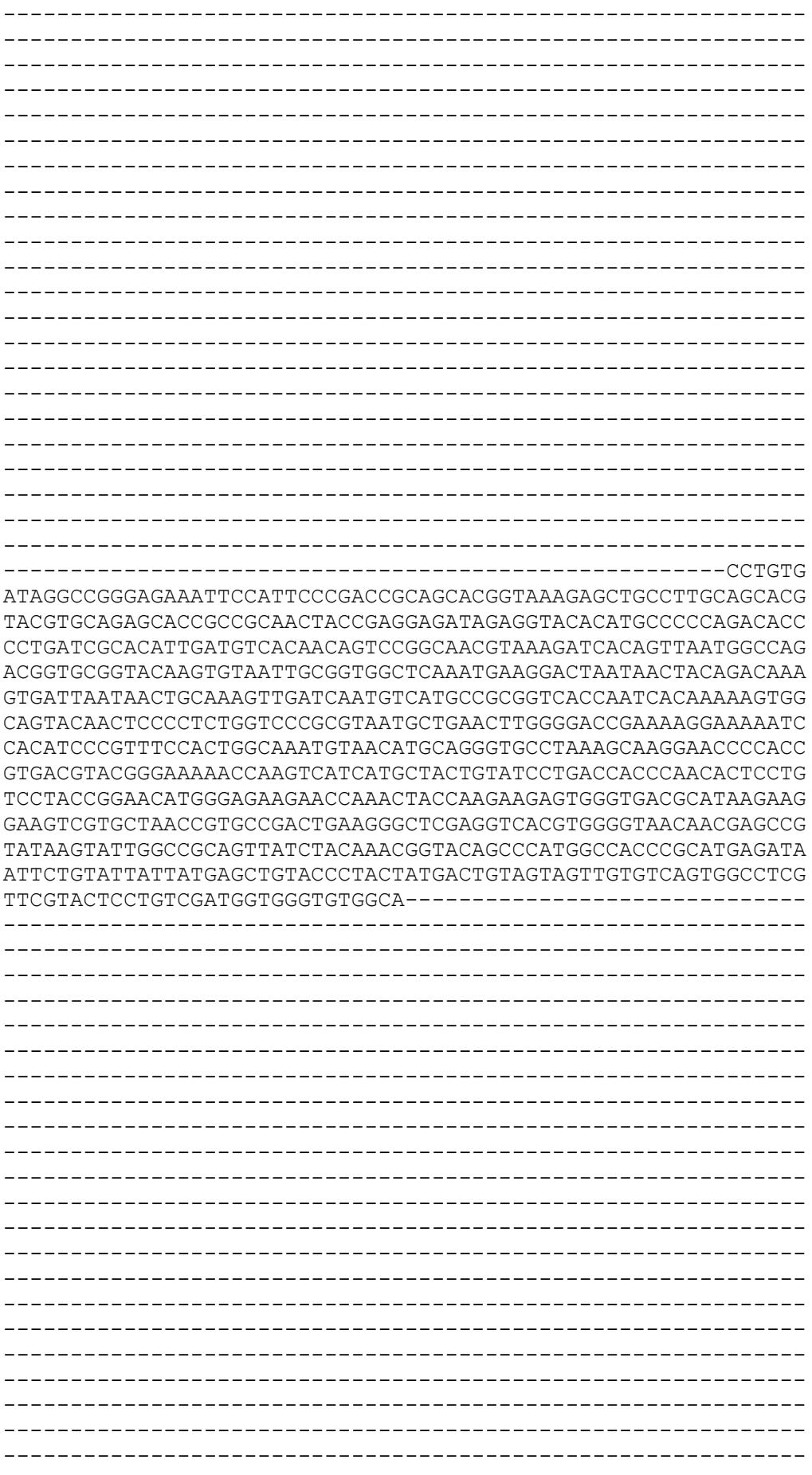
>lcl|MN864720.1_cds_QQL93908.1_1 [protein=E2 glycoprotein]
[protein_id=QQL93908.1] [location=<1..>357] [gbkey=CDS]

--CCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCAAATGAAGGACTAATAACTACAGACAAA
GTGATTAATAACTGCAAAGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGG
CAGTACAACTCCCCTCTGGTCCCGCGTAATGCTGAACTTGGGGACCGAAAA---------

>lcl|MN864721.1_cds_QQL93909.1_1 [protein=E2 glycoprotein]
[protein_id=QQL93909.1] [location=<1..>756] [gbkey=CDS]

--CCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCAAATGAAGGACTAATAACTACAGACAAA
GTGATTAATAACTGCAAAGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGG
CAGTACAACTCCCCTCTGGTCCCGCGTAATGCTGAACTTGGGGACCGAAAAGGAAAAATC
CACATCCCGTTTCCGCTGGCAAATGTAACATGCAGGGTGCCTAAAGCAAGGAACCCCACC
GTGACGTACGGGAAAAACCAAGTCATCATGCTACTGTATCCCGACCACCCAACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCA------------------------------

>lcl|MN864722.1_cds_QQL93910.1_1 [protein=E2 glycoprotein] [protein_id=QQL93910.1] [location=<1..>756] [gbkey=CDS]



>lcl|MN864726.1_cds_QQL93914.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93914.1] [location=<1..>654] [gbkey=CDS]

----------ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCTTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC----------

>lcl|MN864727.1_cds_QQL93915.1_1 [protein=E2-E1 glycoprotein] [protein_id=QQL93915.1] [location=<1..>654] [gbkey=CDS]

---ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCTTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864729.1_cds_QQL93917.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93917.1] [location=<1..>654] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------------------------------ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
```

TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAACAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC-------------------
----

>lcl|MN864733.1_cds_QQL93921.1_1 [protein=E2-E1 glycoprotein] [protein_id=QQL93921.1] [location=<1..>654] [gbkey=CDS]

----------ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAACAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC----------

>lcl|MN864735.1_cds_QQL93923.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93923.1] [location=<1..>654] [gbkey=CDS]


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
----------------------------------------------------ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864734.1_cds_QQL93922.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93922.1] [location=<1..>654] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```


---ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------

>lcl|MN864732.1_cds_QQL93920.1_1 [protein=E2-E1 glycoprotein] [protein_id=QQL93920.1] [location=<1..>264] [gbkey=CDS]

---ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCG---

>lcl|MN864731.1_cds_QQL93919.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93919.1] [location=<1..>654] [gbkey=CDS]


```
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC----------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
----------------------------------------------------------
```

>lcl|MN864730.1_cds_QQL93918.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93918.1] [location=<1..>654] [gbkey=CDS]
ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC

>lcl|MN864728.1_cds_QQL93916.1_1 [protein=E2-E1 glycoprotein] [protein_id=QQL93916.1] [location=<1..>654] [gbkey=CDS]

---ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------

>lcl|MN864725.1_cds_QQL93913.1_1 [protein=E2-E1 glycoprotein] [protein_id=QQL93913.1] [location=<1..>654] [gbkey=CDS]

---ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA

TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------

>lcl|MN864724.1_cds_QQL93912.1_1 [protein=E2-E1 glycoprotein] [protein_id=QQL93912.1] [location=<1..>654] [gbkey=CDS]

---ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG

GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------

>lcl|MN864723.1_cds_QQL93911.1_1 [protein=E2-E1 glycoprotein]
[protein_id=QQL93911.1] [location=<1..>654] [gbkey=CDS]


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
----------------------------------------------------ACACTCCTG
TCCTACCGGAACATGGGAGAAGAACCAAACTACCAAGAAGAGTGGGTGACGCATAAGAAG
GAAGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTCACGTGGGGTAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTACAAACGGTACAGCCCATGGCCACCCGCATGAGATA
ATTCTGTATTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCGTACTCCTGTCGATGGTGGGTGTGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAT
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGCTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCGTCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAATAGACCGGGCTACAGTCCC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
----------------------------
>lcl|MN864738.1_cds_QQL93926.1_1 [protein=E3-E2 envelope protein]
[protein_id=QQL93926.1] [location=<1..>330] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```


------------------AGTCACTCATGTACGCACCCATTTCACCATGACCCTCCTGTG
ATAGGCCGGGAGAAATTCCATTCCCGACCGCAGCACGGTAAAGAGCTGCCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTACCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTTAATGGCCAG
ACGGTGCGGTACAAGTGTAATTGCGGTGGCTCAAATGAAGGACTAATAACTACAGACAAA
GTGATTAATAACTGCAAAGTTGATCAATGTCATGCCGCGGTCACCAAT------------

>HM045811_Tanzania_1953(ROSS low psg)
ATGGAGTTCATCCCAACCCAAACTTTTTACAACAGGAGGTACCAGCCTCGACCCTGGACT
CCGCGCCCTACTATCCAAGTCATCAGGCCCAGACCGCGCCCGCAGAGGCAAGCTGGGCAA
CTTGCCCAGCTGATCTCAGCAGTTAATAAACTGACAATGCGCGCGGTACCCCAACAGAAG
CCACGCAGGAATCGGAAGAATAAGAAGCAAAAGCAAAAGCAACAGGCGCCACAAAACAAC
ACAAACCAAAAGAAGCAGCCACCTAAAAAGAAACCGGCTCAAAAGAAAAAGAAGCCGGGC
CGCAGAGAGAGGATGTGCATGAAAATCGAAAATGACTGTATTTTCGAAGTCAAGCACGAA
GGTAAGGTAACAGGTTACGCGTGCTTGGTGGGGGACAAAGTAATGAAACCAGCACACGTA
AAGGGGACCATCGATAACGCGGACCTGGCCAAATTGGCCTTTAAGCGGTCATCTAAGTAC
GACCTTGAATGCGCGCAGATACCCGTGCACATGAAGTCCGACGCTTCGAAGTTCACCCAT
GAGAAACCGGAGGGGTACTACAACTGGCACCACGGAGCAGTACAGTACTCAGGAGGCCGG
TTCACCATCCCTACAGGTGCGGGCAAACCAGGGGACAGCGGTAGACCGATCTTCGACAAC
AAGGGACGCGTGGTGGCCATAGTCTTAGGAGGAGCTAATGAAGGAGCCCGTACAGCCCTC

TCAGTGGTGACCTGGAATAAAGACATTGTCACTAAAATCACCCCTGAGGGAGCCGAAGAG
TGGAGTCTTGCCATCCCAGTTATGTGCCTGTTGGCAAATACCACGTTCCCCTGCTCCCAG
CCCCCTTGCACACCCTGCTGCTACGAAAAGGAACCGGAGGAAACCCTACGCATGCTTGAG
GACAACGTCATGAGACCTGGGTACTATCAGCTGCTACAAGCATCCTTAACATGTTCTCCC
CACCGCCAGCGACGCAGCACCAAGGACAACTTCAATGTCTATAAAGCCACAAGACCATAC
TTAGCTCACTGTCCCGACTGTGGAGAAGGGCACTCGTGCCATAGTCCCGTAGCACTAGAA
CGCATCAGAAATGAAGCGACAGACGGGACGCTGAAAATCCAGGTCTCCTTGCAAATTGGA
ATAAAGACGGATGGTAGCCATGATTGGACCAAGCTGCGTTACATGGACAATCACATGCCA
GCAGACGCAGAGAGGGCCGGGCTATTTGTAAGAACATCAGCACCATGCACGATTACTGGA
ACAATGGGACACTTCATCCTGGCCCGATGTCCGAAAGGAGAAACTCTGACGGTGGGATTC
ACTGACAGTAGGAAGATTAGTCACTCATGTACGCACCCATTTCACCACGACCCTCCTGTG
ATAGGCCGGGAAAAATTTCATTCCCGACCGCAGCACGGTAAAGAGCTACCTTGCAGCACG
TACGTGCAGAGCACCGCCGCAACTGCCGAGGAGATAGAGGTACACATGCCCCCAGACACC
CCTGATCGCACATTGATGTCACAACAGTCCGGCAACGTAAAGATCACAGTCAACGGCCAG
ACGGTGCGGTATAAGTGTAATTGCGGTGGCTCAAATGAAGGACTAATAACTACAGACAAA
GTGATTAATAACTGCAGGGTTGATCAATGTCATGCCGCGGTCACCAATCACAAAAAGTGG
CAGTATAACTCCCCTCTGGTCCCGCGTAACGCTGAACTCGGGGACCGAAAAGGAAAAATT
CACATCCCGTTTCCGCTGGCAAATGTAACATGCAGGGTGCCTAAAGCAAGGAACCCCACC
GTGACGTACGGGAAAAACCAAGTCATCATGCTACTGTATCCTGACCACCCAACACTCCTG
TCCTACCGGAGTATGGGAGAAGAACCAAACTATCAAGAAGAGTGGGTGACGCACAAGAAG
GAGGTCGTGCTAACCGTGCCGACTGAAGGGCTCGAGGTTACGTGGGGCAACAACGAGCCG
TATAAGTATTGGCCGCAGTTATCTGCAAACGGTACAGCCCACGGCCACCCGCATGAGATA
ATCTTGTACTATTATGAGCTGTACCCTACTATGACTGTAGTAGTTGTGTCAGTGGCCTCG
TTCATACTCCTGTCGATGGTGGGTATGGCAGTGGGGATGTGCATGTGTGCACGACGCAGA
TGCATCACACCGTACGAACTGACACCAGGAGCTACCGTCCCTTTCCTGCTTAGCCTAATA
TGCTGCATCAGAACAGCTAAAGCGGCCACATACCAAGAGGCTGCGGTATACCTGTGGAAC
GAGCAGCAACCTTTGTTTTGGCTACAAGCCCTTATTCCGCTGGCAGCCCTGATTGTCCTA
TGCAACTGTCTGAGACTCTTACCATGCTGTTGTAAAACGTTGGCTTTTTTAGCCGTAATG
AGCATCGGTGCCCACACTGTGAGCGCGTACGAACACGTAACAGTGATCCCGAACACGGTG
GGAGTACCGTATAAGACTCTAGTCAACAGACCGGGCTACAGCCCCATGGTACTGGAGATG
GAGCTACTGTCAGTCACTTTGGAGCCAACGCTATCGCTTGATTACATCACGTGCGAGTAC
AAAACCGTCATCCCGTCTCCGTACGTGAAATGCTGCGGTACAGCAGAGTGCAAGGACAAA
AACCTACCTGACTACAGCTGTAAGGTCTTCACCGGCGTCTACCCATTTATGTGGGGCGGC
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACATGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCAAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATCGTGGTGTACAAAGGTGACGTTTACAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACGCCTGAGAGCAAA
GACGTCTATGCTAACACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACACGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGTTAAAAGAACGAGGGGCGTCGCTA
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGATGAACTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCACTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCATGCGAGGTACCAGCCTGCACCCATTCCTCAGAC
TTTGGGGGCGTCGCCATTATTAAGTATGCAGCCAGCAAGAAAGGCAAGTGTGCGGTGCAT
TCGATGACCAACGCCGTCACTATCCGGGAAGCTGAGATAGAAGTTGAAGGGAATTCTCAG
CTGCAAATCTCTTTCTCGACGGCCTTGGCCAGCGCCGAATTCCGCGTACAAGTCTGTTCT
ACACAAGTACACTGTGCAGCCGAGTGCCACCCTCCGAAGGACCACATAGTCAACTACCCG
GCGTCACATACCACCCTCGGGGTCCAGGACATTTCCGCTACGGCGATGTCATGGGTGCAG
AAGATCACGGGAGGTGTGGGACTGGTTGTCGCTGTTGCAGCACTGATTCTAATCGTGGTG
CTATGCGTGTCGTTCAGCAGGCACTAA
>lcl|MN864713.1_cds_QQL93901.1_1 [protein=E1 glycoprotein]
[protein_id=QQL93901.1] [location=<1..>672] [gbkey=CDS]


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCATCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTACA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
```

>lcl|MN864714.1_cds_QQL93902.1_1 [protein=E1 glycoprotein] [protein_id=QQL93902.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTCCCAGCCTGC---------------

>lcl|MN864716.1_cds_QQL93904.1_1 [protein=E1 glycoprotein] [protein_id=QQL93904.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCGTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA

```
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864717.1_cds_QQL93905.1_1 [protein=E1 glycoprotein] [protein_id=QQL93905.1] [location=<1..>672] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```


---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCGTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTACA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864719.1_cds_QQL93907.1_1 [protein=E1 glycoprotein] [protein_id=QQL93907.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864715.1_cds_QQL93903.1_1 [protein=E1 glycoprotein]
[protein_id=QQL93903.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864712.1_cds_QQL93900.1_1 [protein=E1 glycoprotein] [protein_id=QQL93900.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864711.1_cds_QQL93899.1_1 [protein=E1 glycoprotein] [protein_id=QQL93899.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864710.1_cds_QQL93898.1_1 [protein=E1 glycoprotein]
[protein_id=QQL93898.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864709.1_cds_QQL93897.1_1 [protein=E1 glycoprotein]
[protein_id=QQL93897.1] [location=<1..>672] [gbkey=CDS]

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864708.1_cds_QQL93896.1_1 [protein=E1 glycoprotein] [protein_id=QQL93896.1] [location=<1..>672] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
```

```
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
>lcl|MN864707.1_cds_QQL93895.1_1 [protein=E1 glycoprotein]
[protein_id=QQL93895.1] [location=<1..>672] [gbkey=CDS]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```

---------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTAAA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------

>lcl|MN864718.1_cds_QQL93906.1_1 [protein=E1 glycoprotein] [protein_id=QQL93906.1] [location=<1..>672] [gbkey=CDS]


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
----------------------------------GGCGTCTACCCATTTATGTGGGGCGGA
GCCTACTGCTTCTGCGACGCTGAAAACACGCAATTGAGCGAAGCACACGTGGAGAAGTCC
GAATCATGCAAAACAGAATTTGCATCAGCATACAGGGCTCATACCGCATCCGCATCAGCT
AAGCTCCGCGTCCTTTACCAAGGAAATAACATCACTGTAACTGCCTATGCTAACGGCGAC
CATGCCGTCACAGTTAAGGACGCCAAATTCATTGTGGGGCCAATGTCTTCAGCCTGGACA
CCTTTCGACAACAAAATTGTGGTGTACAAAGGTGACGTCTATAACATGGACTACCCGCCC
TTTGGCGCAGGAAGACCAGGACAATTTGGCGATATCCAAAGTCGCACACCTGAGAGTACA
GACGTCTATGCTAATACACAACTGGTACTGCAGAGACCGGCTGCGGGTACGGTACATGTG
CCATACTCTCAGGCACCATCTGGCTTTAAGTATTGGCTAAAAGAACGAGGGGCGTCGCTG
CAGCACACAGCACCATTTGGCTGCCAAATAGCAACAAACCCGGTAAGAGCGGTGAATTGC
GCCGTAGGGAACATGCCCATCTCCATCGACATACCGGATGCGGCCTTCATTAGGGTCGTC
GACGCGCCCTCTTTAACGGACATGTCGTGCGAGGTACCAGCCTGC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------
```

>QQL93928.1 E3-E2 envelope protein, partial [Chikungunya virus]

----------E
DNVMRPGYYQLLQASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALE
RIRNEATDGTLKIQVSLQIGIKTDDSHDWTKLRYMDNHTPADAERAGLFVRTSAPCTITG
TMGHFILTRCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEGHMPPDTPDRTLMSQQSGNVKITVNGQTVR----------

>QQL93927.1 E3-E2 envelope protein, partial [Chikungunya virus]

----------YLAHCPDCGEGHSCHSPVALE
RIRNEATDGTLKIQVSLQIGIKTDDSHDWTKLRYMDNHTPADAERAGLFVRTSAPCTITG
TMGHFILTRCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQST----------

>QQL93924.1 E3-E2 envelope protein, partial [Chikungunya virus]

----------CCYEREPEETLRMLE
DNVMRPGYYQLLQASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALE
RIRNEATDGTLKIQVSLQIGIKTDDSHDWTKLRYMDNHTPADAERAGLFVRTSAPCTITG
TMGHFILTRCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSN--------

>QQL93925.1 E3-E2 envelope protein, partial [Chikungunya virus]

----------TCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALE
RIRNEATDGTLKIQVSLQIGIKTDDSHDWTKLRYMDNHTPADAERAGLFVRTSAPCTITG
TMGHFILTRCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITTDK
VINNCKVDQCHA----------

>QQL93930.1 E3-E2 envelope protein, partial [Chikungunya virus]

----------HSPVALE
RIRNEATDGTLKIQVSLQIGIKTDDSHDWTKLRYMDNHTPADAERAGLFVRTSAPCTITG
TMGHFILTRCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGS---------

>QQL93929.1 E3-E2 envelope protein, partial [Chikungunya virus]

----------LRMLE

DNVMRLGYYQLLQASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALE
RIRNEATDGTLKIQVSLQIGIKTDDSHDWTKLRYMDNHTPADAERAGLFVRTSAPCTITG
TMGHFILTRCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITT--

>QQL93908.1 E2 glycoprotein, partial [Chikungunya virus]

---PVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITTDK
VINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRK----------

>QQL93909.1 E2 glycoprotein, partial [Chikungunya virus]

---PVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITTDK
VINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRKGKIHIPFPLANVTCRVPKARNPT
VTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVA----------


```
------------------------------------------------------------
--------------------------------------------
>QQL93910.1 E2 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------PVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITTDK
VINNCKVDQCHAAVTNHKKWQYNSPLVPRNAELGDRKGKIHIPFPLANVTCRVPKARNPT
VTYGKNQVIMLLYPDHPTLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVA----------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
--------------------------------------------------
>QQL93923.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
--------------------------------------------------
>QQL93922.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
```

YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----

>QQL93921.1 E2-E1 glycoprotein, partial [Chikungunya virus]

-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----

>QQL93920.1 E2-E1 glycoprotein, partial [Chikungunya virus]

-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLS---------------

>QQL93919.1 E2-E1 glycoprotein, partial [Chikungunya virus]


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93918.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93917.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```


>QQL93916.1 E2-E1 glycoprotein, partial [Chikungunya virus]

-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----

>QQL93915.1 E2-E1 glycoprotein, partial [Chikungunya virus]

-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----

>QQL93914.1 E2-E1 glycoprotein, partial [Chikungunya virus]


```
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93913.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93912.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93911.1 E2-E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
```

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------TLLSYRNMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSTNGTAHGHPHEIILYYYELYPTMTVVVVSVASFVLLSMVGVAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSVGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSP-----
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93926.1 E3-E2 envelope protein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
--------------------------SHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATTEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITTDK
VINNCKVDQCHAAVTN-----------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------------
>ADG95932.1 structural polyprotein [Chikungunya virus]
MEFIPTQTFYNRRYQPRPWTPRPTIQVIRPRPRPQRQAGQLAQLISAVNKLTMRAVPQQK
PRRNRKNKKQKQKQQAPQNNTNQKKQPPKKKPAQKKKKPGRRERMCMKIENDCIFEVKHE
GKVTGYACLVGDKVMKPAHVKGTIDNADLAKLAFKRSSKYDLECAQIPVHMKSDASKFTH
EKPEGYYNWHHGAVQYSGGRFTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTAL
SVVTWNKDIVTKITPEGAEEWSLAIPVMCLLANTTFPCSQPPCTPCCYEKEPEETLRMLE
DNVMRPGYYQLLQASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHSPVALE
RIRNEATDGTLKIQVSLQIGIKTDGSHDWTKLRYMDNHMPADAERAGLFVRTSAPCTITG
TMGHFILARCPKGETLTVGFTDSRKISHSCTHPFHHDPPVIGREKFHSRPQHGKELPCST
YVQSTAATAEEIEVHMPPDTPDRTLMSQQSGNVKITVNGQTVRYKCNCGGSNEGLITTDK
VINNCRVDQCHAAVTNHKKWQYNSPLVPRNAELGDRKGKIHIPFPLANVTCRVPKARNPT
VTYGKNQVIMLLYPDHPTLLSYRSMGEEPNYQEEWVTHKKEVVLTVPTEGLEVTWGNNEP
YKYWPQLSANGTAHGHPHEIILYYYELYPTMTVVVVSVASFILLSMVGMAVGMCMCARRR
CITPYELTPGATVPFLLSLICCIRTAKAATYQEAAVYLWNEQQPLFWLQALIPLAALIVL
CNCLRLLPCCCKTLAFLAVMSIGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSPMVLEM
ELLSVTLEPTLSLDYITCEYKTVIPSPYVKCCGTAECKDKNLPDYSCKVFTGVYPFMWGG
```

AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAMNC
AVGNMPISIDIPDAAFTRVVDAPSLTDMSCEVPACTHSSDFGGVAIIKYAASKKGKCAVH
SMTNAVTIREAEIEVEGNSQLQISFSTALASAEFRVQVCSTQVHCAAECHPPKDHIVNYP
ASHTTLGVQDISATAMSWVQKITGGVGLVVAVAALILIVVLCVSFSRH
>QQL93907.1 E1 glycoprotein, partial [Chikungunya virus]

----------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC----------

>QQL93904.1 E1 glycoprotein, partial [Chikungunya virus]

----------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC----------

>QQL93903.1 E1 glycoprotein, partial [Chikungunya virus]


```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC-------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93902.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC-------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93900.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC-------------------------
------------------------------------------------------------
-------------------------------------------------
```

>QQL93899.1 E1 glycoprotein, partial [Chikungunya virus]

---GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC-------------------------

>QQL93898.1 E1 glycoprotein, partial [Chikungunya virus]

---GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC-------------------------

>QQL93897.1 E1 glycoprotein, partial [Chikungunya virus]


```
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93896.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93895.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESK
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC------------------------
------------------------------------------------------------
-------------------------------------------------
>QQL93906.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPEST
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC------------------------
------------------------------------------------------------
--------------------------------------------------
>QQL93905.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPEST
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC------------------------
------------------------------------------------------------
--------------------------------------------------
>QQL93901.1 E1 glycoprotein, partial [Chikungunya virus]
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------------GVYPFMWGG
AYCFCDAENTQLSEAHVEKSESCKTEFASAYRAHTASASAKLRVLYQGNNITVTAYANGD
HAVTVKDAKFIVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPEST
DVYANTQLVLQRPAAGTVHVPYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNC
AVGNMPISIDIPDAAFIRVVDAPSLTDMSCEVPAC------------------------
```

Supplementary Table 1. Primers used to amplify the genes of the structural envelope polyproteins. The region amplified by the primers for each envelope protein gene is indicated (target region).

Primer	Target region	Sequence 5'-3'	Amplicon (bp)
P1_Foward	E1: 132-424	CAGCAGAGTGCAAGGACAAA	882
P1_Reverse		CCGTCGAGAAAGAGATTTGC	
P2_Forward	E1: 251-403	AACATGCAGGGTGCCTAAAG	855
P2_Reverse	E2: 1-137	GTACCGCAGCATTTCACGTA	
P3_Foward	E2: 93-393	ACTCTGACGGTGGGATTCAC	891
P3_Reverse		AGCTCCTGGTGTCAGTTCGT	
P4_Forward	E2: 1-236	AGTCTTGCCATCCCAGTCAT	943
P4_Reverse	E3: 6-64	GGTCCCCAAGTTCAGCATTA	

Supplementary Table 2. Representative sequences from CHIKV genotypes used in the phylogenetic analysis, indicating the GenBank accession number, geographical location and year of virus isolation.

GenBank	Country	Year of isolation
HM045811	Tanzania	1953
LC259082	Thailand	1958
HM045809	Democratic Republic of Congo	1960
HM045823	Angola	1962
HM045803	India	1963
HM045786	Nigeria	1964
HM045788	India	1973
HM045814	Thailand	1975
HM054822	Central African Republic	1978
HM045812	Uganda	1982
HM045791	Indonesia	1983
AY726732	Senegal	1983
HM045784	Central African Republic	1984
HM04597	Indonesia	1985
HM045790	Philippines	1985
HM045789	Thailand	1988

GenBank	Country	Year of isolation
HM045786	Thailand	1995
MH229986	Mauritius Islands	2006
MK098629	Sri Lanka	2006
EU703759	Malaysia	2006
FJ807897	Indonesia	2007
Kt324224	Malaysia	2008
FJ445510	Singapore	2008
Lc259093	Malaysia	2009
MH124573	India	2010
MF773565	East Timor	2010
MF773567	Borneo	2011
HE806461	New Caledonia	2011
MF773561	Bali	2011
MG664850	China	2012
MG664851	China	2012
LC259084	Filipinas	2012
MF773569	Papua New Guinea	2013

GenBank	Country	Year of isolation
KJ451622	Micronesia	2013
KP164568	Feira de Santana-BA/Brazil	2014
KX629422	India	2014
LC259088	Tonga	2014
MF773559	Samoa	2014
KP164571	Recife-PE/Brazil	2014
MF773560	Caribbean	2014
MH359139	Colombia	2014
KP164572	Belém do Pará-PA/Brazil	2014
KJ451624	British Virgin Islands	2014
LC259089	Dominica	2014
MH329299	Colombia	2014
KP164567	Oiapoque-AP	2014
MG967666	Haiti	2014
KR559471	El Salvador	2014
KU940225	Salvador-BA/Brazil	2015
MF773562	Kiribati	2015

GenBank	Country	Year of isolation
MG921596	Mexico	2015
KU355832	Rio de Janeiro-RJ/Brazil	2015
LC259090	Colombia	2015
MG000876	Haiti	2016
KX228391	Recife-PE/Brazil	2016
Ky124328	Rio de Janeiro-RJ/Brazil	2016
KY055011	Aracaju-SE/Brazil	2016
MK370032	India	2016
MF774614	Pakistan	2016
MH124581	India	2016
MF499120	Hong Kong	2016
LC259094	Angola	2016
LC259092	Cuba	2016
AF079456	Uganda (Outside group: O'nyong nyong)	1996-1997