Journal of General Virology (2002), **83**, 3075–3084. Printed in Great Britain

Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site

Afjal Hossain Khan, Kouichi Morita, Maria del Carmen Parquet, Futoshi Hasebe, Edward G. M. Mathenge and Akira Igarashi†

Department of Virology, Institute of Tropical Medicine, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

In this study, the complete genomic sequence of chikungunya virus (CHIK; S27 African prototype) was determined and the presence of an internal polyadenylation [I-poly(A)] site was confirmed within the 3′ non-translated region (NTR) of this strain. The complete genome was 11805 nucleotides in length, excluding the 5′ cap nucleotide, an I-poly(A) tract and the 3′ poly(A) tail. It comprised two long open reading frames that encoded the non-structural (2474 amino acids) and structural polyproteins (1244 amino acids). The genetic location of the non-structural and structural proteins was predicted by comparing the deduced amino acid sequences with the known cleavage sites of other alphaviruses, located at the C-terminal region of their virus-encoded proteins. In addition, predicted secondary structures were identified within the 5′ NTR and repeated sequence elements (RSEs) within the 3′ NTR. Amino acid sequence homologies, phylogenetic analysis of non-structural and structural proteins and characteristic RSEs revealed that although CHIK is closely related to o'nyong-nyong virus, it is in fact a distinct virus. The existence of I-poly(A) fragments with different lengths (e.g. 19, 36, 43, 91, 94 and 106 adenine nucleotides) at identical initiation positions for each clone strongly suggests that the polymerase of the alphaviruses has a capacity to create poly(A) by a template-dependant mechanism such as 'polymerase slippage', as has been reported for vesicular stomatitis virus.

Introduction

Chikungunya virus (CHIK) is a member of the *Alphavirus* genus of the family *Togaviridae*. The alphaviruses consist of 30 species of arthropod-borne viruses, which are further sub-grouped into seven serocomplexes based on serological data (Porterfield, 1980; Strauss & Strauss, 1994; Van Regenmortel *et al.*, 2000). CHIK was first isolated from the serum of a febrile patient during a dengue epidemic that occurred in the Newala district, Tanzania, in 1953 (Ross, 1956).

The alphaviruses are enveloped particles and their genome consists of a single-stranded, positive-sense RNA molecule of approximately 12000 nucleotides. The 5′ end is capped with a 7-methylguanosine while the 3′ end is polyadenylated. The non-structural proteins are translated directly from the 5′ two-thirds of the genomic RNA. A subgenomic positive-strand RNA referred to as 26S RNA, identical to the 3′ one-third of the genomic RNA, is transcribed from a negative-stranded RNA intermediate. This RNA serves as the mRNA for the synthesis of the viral structural proteins (Strauss & Strauss, 1986, 1988; Faragher *et al.*, 1988). According to the genomic organization of other alphaviruses, the genome of CHIK is considered to be: 5′ cap-nsP1-nsP2-nsP3-nsP4-(junction region)-C-E3-E2-6K-E1-poly(A) 3′.

Alphaviruses possess conserved sequences at the 5′ and 3′ ends as well as the intergenic region. Conserved repeated sequence elements (RSEs) are also present in the 3′ non-translated region (NTR) among alphaviruses. These conserved domains play an important role in the regulation of viral RNA synthesis (Ou *et al.*, 1981, 1982a, b, 1983; Pfeffer *et al.*, 1998).

CHIK is an important human pathogen that causes a disease syndrome characterized by fever, headache, rash, nausea, vomiting, myalgia and arthralgia (Thaikruea *et al.*, 1997; Diallo *et al.*, 1999; Powers *et al.*, 2000). Its association with a fatal haemorrhagic condition was reported in India (Sarkar *et al.*, 1964). CHIK is geographically distributed from Africa through

Author for correspondence: Kouichi Morita.
Fax +81 95 849 7830. email moritak@net.nagasaki-u.ac.jp

†**Present address:** Kyushu University of Health and Welfare, Yoshinomachi 1714-1, Nobeoka City, Miyazaki 882-8508, Japan.

The GenBank accession number of the sequence reported in this paper is AF369024.

Southeast Asia and South America, and its transmission to humans is mainly through *Aedes* species mosquitoes (Diallo *et al.*, 1999). CHIK activity in Asia has been documented since its isolation in Bangkok, Thailand, in 1958 (Hammon *et al.*, 1960).

O'nyong-nyong virus (ONN) is considered to be a subtype of CHIK. This is because serological tests reveal a one-way antigenic cross-reactivity between the two agents (Chanas *et al.*, 1979; Calisher *et al.*, 1980; Blackburn *et al.*, 1995). However, Powers *et al.* (2000) reported that CHIK and ONN were two distinct viruses after phylogenetic analysis (E1 protein) and serological studies.

Although the 26S sequence of the CHIK genome (vaccine and Ross strains) is available in GenBank (accession nos L37661 and AF490259), the complete nucleotide sequence of a CHIK strain is not available. In the present study, the complete nucleotide sequence of the CHIK genome (strain S27 African prototype) was determined, and the homology of the nucleotide and amino acid sequences and the structure of the viral RNA of CHIK were precisely compared with other alphaviruses, with particular emphasis on the relationship with ONN.

Methods

■ **Virus propagation.** CHIK was inoculated into a monolayer culture of an *Aedes albopictus* clone C6/36 cell line (Igarashi, 1978) and incubated at 28 °C in Eagle's minimum essential medium supplemented with 2% heat-inactivated foetal calf serum and 0·2 mM non-essential amino acids. The infected culture fluid was harvested 4–5 days after inoculation and centrifuged. The supernatant was then filtered with a 0·22 μm filter unit (Millex-GV) and the virus concentrated by ultracentrifugation (Optima L-90K Ultracentrifuge; Beckman) at 175 000 ***g*** for 3 h using an SW 41 rotor. The pellet obtained was dissolved in STE buffer (0·1 M NaCl, 0·01 M Tris–HCl, pH 7·4, 0·001 M EDTA) and kept at −80 °C until use.

■ **RNA extraction and RT–PCR.** RNA was extracted from previously stored virus using Trizol LS (Gibco BRL) following the manufacturer's instructions. To obtain short PCR products up to 1·2 kb in length, RT–PCR was performed as previously described by Morita *et al.* (1991).

To obtain long PCR products above 3·0 kb in length, cDNA was synthesized using Rever Tra Ace (MMLV reverse transcriptase RNaseH$^-$; Toyobo). A 50 μl reaction volume was prepared as follows: a pre-mix consisting of 2·5 μl reverse primer (50 pmol/ μl), 5·0 μl 10 mM dNTP mix (Gibco BRL), 10·0 μl 5 × RT buffer (Toyobo), 2·5 μl Prime RNase inhibitor (30 units; Eppendorf) and 27·5 μl autoclaved dH_2O was prepared and kept at room temperature for 30 min to eliminate any trace of RNase activity. This pre-reaction mixture (47·5 μl) was transferred to an Eppendorf tube containing the extracted RNA pellet and mixed, and finally 2·5 μl of Rever Tra Ace (100 U/ μl) was added. After careful mixing, the RT reaction mixture (50 μl) was incubated at 55 °C for 1 h and then heated to 85 °C for 5 min to inactivate the enzyme. To remove the RNA complementary to the cDNA product, 5 units ribonuclease H (Takara) was added to the reaction mixture and incubated for 30 min at 37 °C. The long PCR amplification was performed using LA *Taq* polymerase (Takara) according to the manufacturer's instructions, applying the following thermal cycling conditions: 94 °C for 3 min, followed by 20 cycles of 94 °C for 30 s and 68 °C for 4 min (1 kb per min) and a final extension at 72 °C for 10 min. The amplified PCR products were analysed by electrophoresis on 1–3% (w/v) agarose gel in TAE buffer followed by brief staining with ethidium bromide. The amplified products were visualized under UV light, excised from the gel and purified with the QIAEX-II Gel Extraction Kit (Qiagen) following the manufacturer's instructions. The purified PCR products were then used for direct sequencing.

■ **Direct sequencing of 5′ and 3′ ends.** The 3′ terminal sequence was determined using a 3′ RACE Kit (Gibco BRL) following the manufacturer's instructions and the 5′ terminal region was sequenced according to the CapSite cDNA manual (Nippon Gene Co.) with slight modifications. Briefly, to remove the cap structure, viral RNA was first dephosphorylated with calf intestinal alkaline phosphatase (Gibco BRL) and the dephosphorylated RNA was then incubated with tobacco acid pyrophosphatase (Nippon Gene Co.). The decapped RNA was ligated to a 30-mer oligoribonucleotide (5′ AGCAUCGAGUCGGCCUUGUUG-GCCUACUGG 3′) using T4 RNA ligase (Takara). The ligated RNA was then subjected to RT–PCR as previously described by Morita *et al.* (1991) using a primer complementary to the nsP1 gene (225C/CHIK nsP1; Table 1) and a 21-mer DNA forward primer (RC primer, 5′ AGCATCG-AGTCGGCCTTGTTG 3′) which had the exact nucleotide sequence of the oligoribonucleotide added at the 5′ end of the viral RNA and which was specific for recapping. Purification of the PCR products was carried out using the methodology described above.

■ **Conventional cDNA cloning to confirm the internal polyadenylation site within the 3′ NTR.** Single-stranded cDNA was synthesized using a reverse primer (11734C/CHIK; Table 1) as described above. The cDNA was extracted with phenol–chloroform, and then washed three times with Microcon YM-100 (Millipore) using autoclaved dH_2O and collected in 50 μl autoclaved dH_2O. Twenty-five μl of this single-stranded cDNA was then mixed with 75 μl of second-strand reaction solution composed of 5 μl sense primer (100 pmol), 11300S/CHIK (Table 1), 10 μl 10 × buffer, 10 μl 25 mM $MgCl_2$, 16 μl 2·5 mM dNTP mix, 1 μl LA *Taq* polymerase (5 units) and 33 μl autoclaved dH_2O. The reaction mixture was heat-denatured at 95 °C for 1 min and incubated at 72 °C for 30 min. The double-stranded cDNA was purified by phenol–chloroform extraction and washed with Microcon YM-100 as above. The double-stranded cDNA was then digested with *Eco*RI (Takara), ligated to pUC118 *Eco*RI/BAP (Takara) and used to transform competent NovaBlue cells (Novagen).

■ **Sequencing strategy.** Nucleotide sequencing was done by the primer extension dideoxy chain termination method (Fig. 1a). For each sequencing reaction, 30–90 ng purified PCR product was combined with 3·2 pmol primer and dRhodamine Terminator Cycle Sequencing Ready Reaction Mixture containing the four dye-labelled deoxynucleotide terminators (Perkin-Elmer/Applied Biosystems). The thermal cycle sequencing parameters used were as described by the manufacturer. The reaction mixture was column-purified (Centri-Sep) and the DNA vacuum-dried for 25–30 min. The pellet was then resuspended in 15 μl of template suppression reagent, heated at 92 °C for 2 min and kept on ice until loaded into the sequencer (ABI Prism 310 Genetic Analyser, Perkin-Elmer/Applied Biosystems). RT–PCR primers were designed using previously published partial sequences for CHIK and the conserved regions of other alphaviruses (Table 1). For nucleotide sequencing, internal primers were designed from the nucleotide sequences of CHIK analysed in this study.

■ **Phylogenetic analysis.** Relationships among the aligned amino acid sequences of the alphaviruses were determined using the PHYLIP program, version 3.5c (Felsenstein, 1985, 1993). The Kimura two-parameter algorithm was applied to calculate the evolutionary distances and the neighbour-joining method was used to construct the phylogram, which was viewed using the TREEVIEW program (Page, 1996). Bootstrap

Table 1. Oligonucleotide primers used for RT–PCR

Primer name and location	Sequence (5′ → 3′)	Accession no.*
41S/CHIK nsP1	TGGATATTGGTAGTGCGCCAGCAAGGAGGATGATGTCGGACAG	U94597
180S/CHIK str. pt.	GCCGCGCAAGAATCGGAAGAATAAGAAGCAAAAGCAAAAGCAGCA	L37661
230C/CHIK str. pt.	GGCGCCTGCTGCTTTTGCTTTTGCTTCTTATTCTTCCGATTCTTG	L37661
3420C/CHIK str. pt.	ATGCACCGCACACTTGCCTTTCTTGCTGGCTGCATATTTAATGAT	L37661
3737S/ONN	AATATCCATACTCCATTCCGCATACATCATTACCAGCAGTGTG	M20303
3998C/ONN	CATTGTCGAATCGGCTAAATAGGAAGAACATTTCTGTATTGCT	M20303
32S/5′ ntr ONN	GGTTTCATACTGCTCTACTC	M20303
203C/CHIK nsP1†	GGCGCACTACCAATATCCA	AF369024
225C/CHIK nsP1†	GTCCGACATCATCCTCCTT	AF369024
444C/CHIK nsP1†	CGCGACGTCTGCTCTCTG	AF369024
11300S/CHIK†	AAAGAATTCAGCAGGCACTAACTTGAC	AF369024
11734C/CHIK†	AAAGAATTCTGAGTTCGGCTGCTTTTAGG	AF369024
RC‡	AGCATCGAGTCGGCCTTGTTG	
3300S/CHIK str. pt.	CGACGCGCCATCTTTAAC	L37661
3708S/CHIK str. pt.	CCTAATCGTGGTGCTATG	L37661
217S/3′ ntr CHIK	CAACCCCTGAATAGTAACAA	AF023283
413C/3′ ntr CHIK	TGAGTTCGGCTGCTTTTAGG	AF023283
459C/3′ ntr CHIK	CGGAGAATCGTGGAAGAGT	AF023283
AP§	GGCCACGCGTCGACTAGTAC$(T)_{17}$	
AUAP§	GGCCACGCGTCGACTAGTAC	

* Accession number of the viruses from which primers were designed.
† Primers designed from the sequence presented in this study.
‡ A primer specific for recapping oligoribonucleotide (5′ AGCAUCGAGUCGGCCUUGUUGGCCUACUGG 3′) to amplify the 5′ end.
§ Oligonucleotides provided within the 3′ RACE kit used in this study.
S or C in a primer name indicates sense and complementary; str. pt., structural protein; underlined, additional external sequence included to create *Eco*RI site for conventional cDNA cloning.

analysis was performed on 1000 replicas using the programs SEQBOOT and CONSENSE to ascertain support for the major branches of the tree. The GenBank accession numbers for the sequences of alphaviruses used in this study are ONN (M20303), Barmah Forest virus (BF; U73745), Semliki Forest virus (SF; X04129), Ross River virus (RR; M20162, K00046), eastern equine encephalitis virus (EEE; S26372, X63135), western equine encephalitis virus (WEE; AF143811), Venezuelan equine encephalitis virus (VEE; L04653), Sindbis virus (SIN; J02363) and CHIK (AF023283, AF192895, AF192907, L37661 and U94597). The computer analysis of nucleotide and deduced amino acid sequences was carried out using the DNASIS Mac version 3.6 Software System (Hitachi, Japan, 1995) and BLAST (Altschul *et al.*, 1990).

Results

Complete nucleotide sequence of the CHIK genome

The nucleotide sequence of CHIK was determined bi-directionally from cDNA by directly sequencing RT–PCR products, according to the sequencing strategy summarized in Fig. 1(a). The genomic RNA was 11805 nucleotides long, excluding the 5′ cap nucleotide, an I-poly(A) site and the 3′ poly(A) tract. The 5′ NTR was composed of 76 nucleotides, the 3′ NTR of 526 nucleotides and the junction region, which is also untranslated, of 68 nucleotides. The calculated base composition was as follows: 30% A, 25% C, 25% G and 20% T(U).

The non-structural proteins were contained in an open reading frame of 7425 nucleotides initiated by a start codon triplet (ATG) at position 77–79 and terminated at a stop codon triplet (TAG) at position 7499–7501. This open reading frame encoded a polyprotein of 2474 amino acids from which the individual non-structural proteins are formed by proteolytic cleavage. It was confirmed by us that CHIK, as in ONN and SF, has a sense codon (CGA encoding Arg) at the C-terminal region of nsP3 in place of the opal (TGA) termination codon present in other alphaviruses (RR, SIN, WEE, EEE, VEE and BF).

The subgenomic RNA (26S RNA) was 4327 nucleotides long excluding only the poly(A) tract at the 3′ end and started at position 7498. The structural proteins were contained in an open reading frame of 3735 nucleotides initiated by a start codon at position 7567–7569 and terminated by a stop codon at position 11299–11301. This open reading frame encodes a polyprotein of 1244 amino acids from which the individual structural proteins are formed.

Characteristics of CHIK non-structural proteins

The nsP1 was 535 amino acids long. Seventeen amino acids (QVTPNDHANARAFSHLA) conserved among alphaviruses were identified near the N terminus of nsP1 at position 31–47.

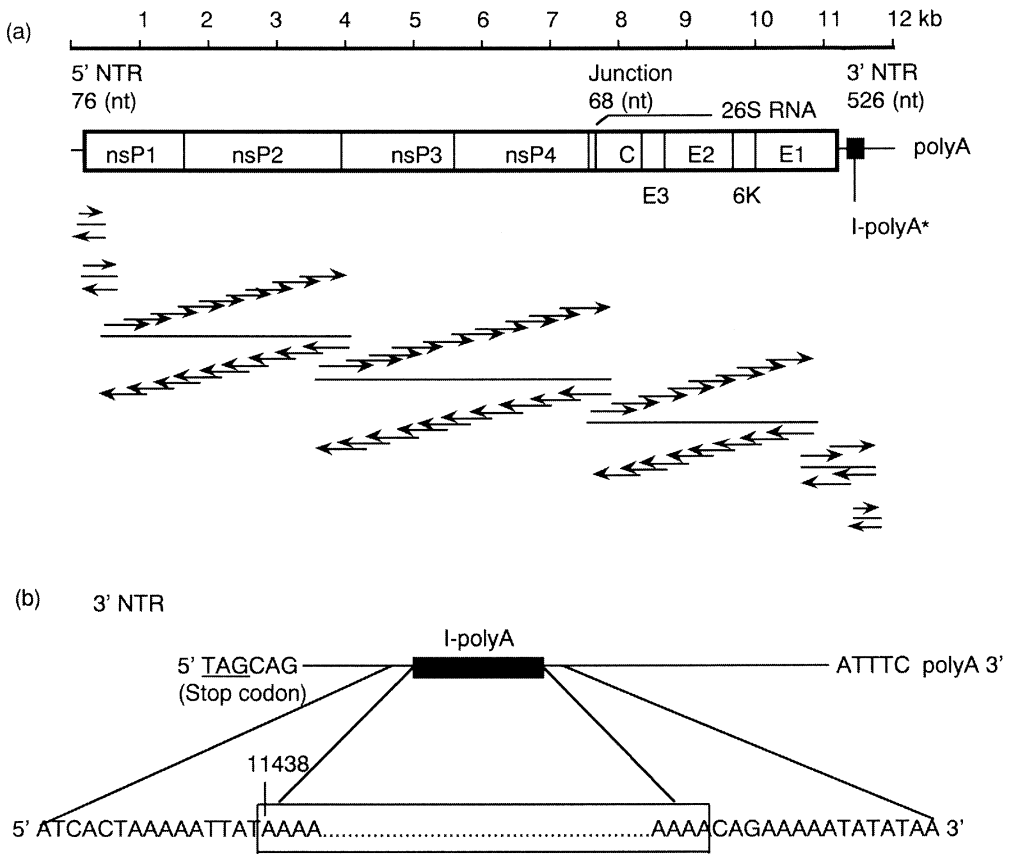

Fig. 1. (a) Sequencing strategy for CHIK genomic RNA. The scale is indicated in kilobases. Boxes indicate the coding regions of non-structural and structural proteins. The location of the subgenomic RNA is indicated by 26S RNA. The horizontal lines at the two ends indicate the non-coding regions. The whole genome was amplified by RT–PCR for bi-directional sequence analysis. *I-poly(A), internal poly(A) site. (b) Structure around the I-poly(A) site. The length of the I-poly(A) tract varies as shown in Table 3.

Table 2. Percentage identities among non-structural and structural polyproteins of alphaviruses

The percentage identity was determined from pairwise comparisons of amino acid sequences of the non-structural polyprotein (top right of table) and the structural polyprotein (bottom left of table) of alphaviruses.

	CHIK	CHIK*	EEE	VEE	WEE	RR	SF	BF	ONN	SIN
CHIK		NA	58	58	58	68	70	59	85	59
CHIK*	96		NA	NA	NA	NA	NA	NA	NA	NA
EEE	48	47		67	79	58	59	57	58	56
VEE	45	46	56		67	59	58	57	58	56
WEE	43	42	56	50		58	59	56	58	56
RR	60	60	47	45	45		72	62	67	57
SF	62	62	48	45	45	73		60	68	59
BF	50	50	47	44	44	54	55		63	56
ONN	85	84	47	46	43	59	60	51		58
SIN	44	44	49	46	68	47	46	45	43	

* The 26S sequence of CHIK vaccine strain was included for analysis.
NA: the sequence for the non-structural polyprotein of the CHIK vaccine strain was not available.

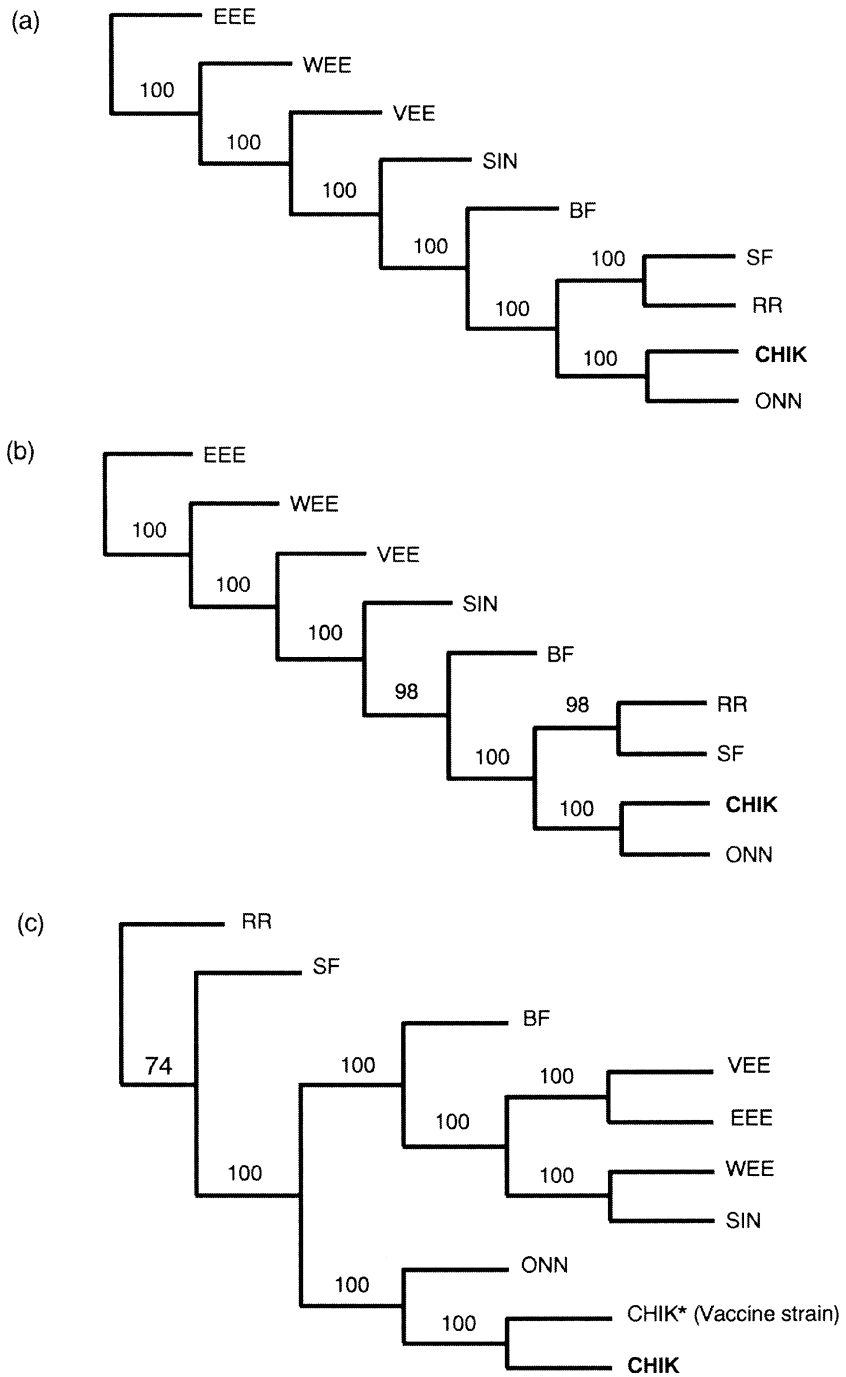

Fig. 2. Phylogenetic relationship of CHIK to other alphaviruses. The phylograms for the predicted amino acid sequences of the complete coding region (a), the non-structural polyprotein (b) and the structural polyprotein (c) are shown. The bootstrap values for each branch of the trees are indicated for 1000 replicas. *, The 26S sequence of Chikungunya virus vaccine strain was included for analysis.

The nsP2, the largest non-structural protein among alphaviruses, was 798 amino acids long in the CHIK strain used in this study. CHIK contained a large net positive charge (+21) in this protein, similar to other alphaviruses. The replicase motif (GXXXXGKS, where X represents any amino acid) was found in the nsP2 of CHIK at position 186–193. A three amino acid motif (CWA) of the non-structural proteinase among alphaviruses was also identified in the nsP2 of CHIK at position 478–480. The degree of identity between CHIK and the other alphaviruses for the deduced nsP2 amino acid sequence ranged from 56% (WEE) to 92% (ONN) (data not shown). The nsP3 for CHIK was 530 amino acids long. This protein has a large

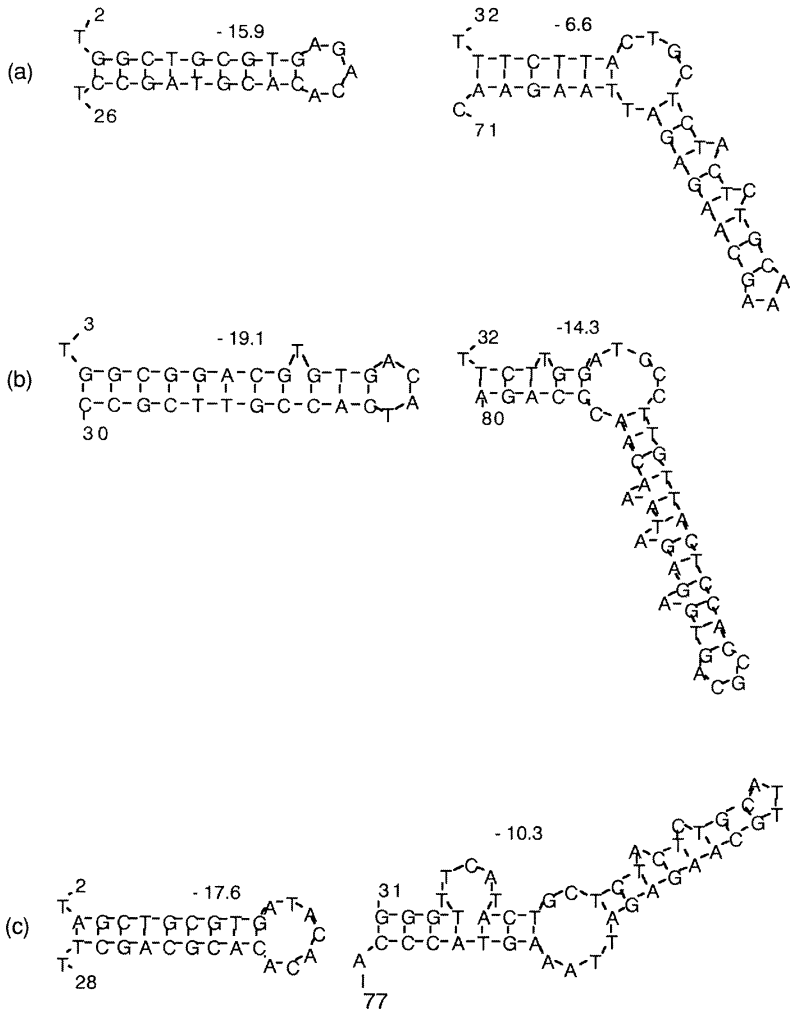

Fig. 3. Secondary structures of the 5′ NTR. Stem–loop structures were predicted in the 5′ NTR of CHIK (a), RR (b) and ONN (c) using the DNASIS Mac version 3.6 software system (Hitachi, Japan, 1995). The minimum free energy values (kcal/mol) for the different structures are shown.

net negative charge in CHIK and ONN (−24 and −25, respectively) compared with the net negative charges recorded for RR, SF and SIN (−12, −10 and −8, respectively). The nsP4 contained 611 amino acids. The deduced amino acid sequence identity between CHIK and the other alphaviruses for the nsP4 ranged from 71% (BF) to 91% (ONN) (data not shown) indicating that it is the best-conserved protein among the alphaviruses. The motif Gly-Asp-Asp (GDD) of the RNA polymerase was found to be located at position 465–467, near the C terminus of the nsP4 sequence for the CHIK isolate presented herein.

Characteristics of CHIK structural proteins

The capsid (C) protein was 261 amino acids long and the E3 protein of CHIK consisted of 64 amino acids. The E2 protein had two possible glycosylation sites at positions 263 and 345 assigned by the sequence Asn-X-Ser/Thr (where X is any amino acid except proline) and it contained 423 amino acids. In this protein, CHIK shared an 82% amino acid sequence identity with ONN. The total length of the 6K protein was 61 amino acids. The E1 protein contained 435 amino acids, and a possible glycosylation site was identified at position 141. In this region, the amino acid sequence identity between CHIK and ONN was 88%. The E1 protein of CHIK contained an uncharged tract (residues 80–96).

Percentage identities of non-structural and structural polyproteins among alphaviruses

The percentage of amino acid sequence identity between CHIK and the other alphaviruses was determined using BLAST (Altschul *et al.*, 1990) for the non-structural and structural polyproteins (Table 2). For the non-structural polyprotein, the degree of identity between CHIK and the other alphaviruses ranged from 58% to 85%, with ONN as the closest related virus and EEE, VEE, WEE, BF and SIN the most distantly related viruses. For the structural polyprotein, however, the degree of identity between CHIK and the other alphaviruses was wider, ranging from 42% to 85%. The CHIK structural polyprotein had 85% identity with that of ONN, but only 42% identity with that of WEE. A comparison of amino acid sequences from the C-terminal regions of the viral-encoded proteins among alphaviruses was made to predict the cleavage

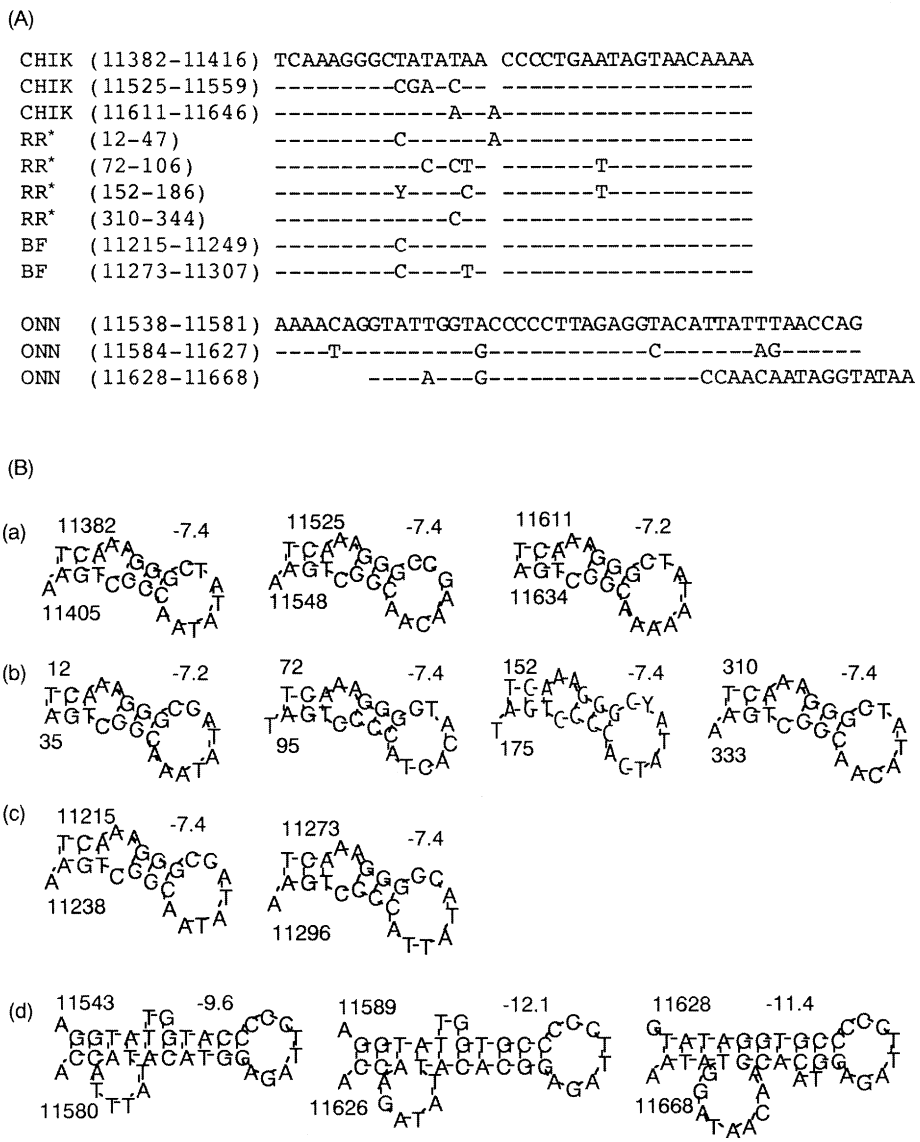

Fig. 4. (A) Comparison of RSEs among several alphaviruses. The dashes indicate that the nucleotide is identical to that found in the CHIK sequence. The RSEs in the 3′ NTR of CHIK, RR, BF and ONN were compared. ONN contained two complete and one incomplete RSEs, which are different from CHIK and the other alphaviruses presented. *, Positions are indicated from the nucleotide sequence of the 3′ NTR of RR. (B) Predicted secondary structures of RSEs in the 3′ NTR. Presented are the hairpin structures of the RSEs in the 3′ NTR of CHIK (a), RR (b), BF (c) and ONN (d). The minimum free energy values (kcal/mol) are shown for each structure.

sites of CHIK generated by proteolytic activity, as described in Strauss & Strauss (1994).

Phylogenetic analysis

To date, there are very few reports on the molecular relationship between CHIK and the other members of the *Alphavirus* genus, and only available partial sequences of CHIK were considered for analysis (Lee *et al.*, 1997; Lanciotti *et al.*, 1998; Powers *et al.*, 2000). In this report, the amino acid sequence of the complete coding region of CHIK was subjected to phylogenetic analysis using the neighbour-joining method. Using either the complete coding region or the amino acid sequences of the non-structural and structural polyproteins independently, the phylogenetic trees generated indicated that the closest-related virus to CHIK is ONN, followed by RR, SF and BF. The most distantly related viruses are EEE, WEE, VEE and SIN (Fig. 2).

Non-translated regions

Ou *et al.* (1983) reported that the nucleotide sequences at the 5′ termini of alphaviruses are conserved more in potential secondary structure than in sequence and this conserved secondary structure may be of importance for virus RNA replication. The secondary structures were predicted for the 5′

Table 3. Length of the internal poly(A) in the 3′ NTR of the CHIK isolate used in this study, determined by conventional cDNA cloning methods

Clone no.	Length of I-p(A)
Clone 1	19*
Clone 2	36
Clone 3	43
Clone 4	91
Clone 5	94
Clone 6	106

* Clone 1, which included an I-poly(A) fragment of 19 adenine nucleotides, corresponds to GenBank accession no. AF369024.

NTRs of alphaviruses (Fig. 3). The first stem–loop structure identified at the 5′ terminus (nt 2–26) of CHIK was identical to those identified for RR and ONN. The second stem–loop structure (nt 32–71) was only identical to that found for RR. ONN, in contrast, possessed a separate loop in its second stem–loop structure and the angulation was different.

Alphaviruses contain repeated sequence elements (RSEs) in the 3′ NTR (Pfeffer *et al.*, 1998). The 3′ NTR of CHIK contained three RSEs at positions 11382–11416, 11525–11559 and 11611–11646. Four nucleotide differences and one nucleotide difference and one insertion were observed for the second and third RSEs, respectively. RR and BF contained four and two RSEs found for CHIK, respectively. Secondary structures were predicted using the first 24 nucleotides of each RSE. All of the predicted hairpin structures were identical (Fig. 4A, B). Even though ONN is closely related to CHIK, it did not contain any of the RSEs found for CHIK. The CHIK isolate used in this study as well as the ONN (Levinson *et al.*, 1990) manifested different types of RSE and secondary structures (Fig. 4A, B). Within the 3′ NTR, the CHIK genome contained 19 nucleotides (5′ ATTTTGTTTTTAATATTTC 3′) adjacent to the poly(A) tail, which have been shown to be conserved among alphaviruses.

Identification of an internal poly(A) tract

An I-poly(A) tract within the 3′ NTR of the CHIK genome, beginning at nt 11438, was found by direct sequencing of the RT–PCR product. It was not possible to determine the length of the I-poly(A) by this method. Cloned PCR products in pCR 2·1 plasmid possessed different lengths of the I-poly(A) (data not shown). In order to exclude the possibility that the I-poly(A) was generated by PCR, cDNA from this region was cloned by conventional cDNA cloning without using PCR, as described in the Methods section. Six clones were sequenced. The length of the I-poly(A) fragment varied from 19 to 106 adenine nucleotides, as shown in Table 3, suggesting that it was not a PCR artefact and that this polyadenylation occurred in different viral RNA syntheses. Clone 1, which included an I-poly(A) fragment of 19 adenine nucleotides, corresponded to the GenBank accession number AF369024.

Discussion

We completed the entire nucleotide sequence analysis of the African prototype of Chikungunya virus (S27 strain) and compared the nucleotide and deduced amino acid sequences with those of other alphaviruses, with particular attention to ONN.

The non-structural proteins among alphaviruses contain a number of common characteristics, such as 17 consensus amino acids at the N terminus of nsP1 (Ou *et al.*, 1983), a large net positive charge for nsP2 (Strauss *et al.*, 1984; Faragher *et al.*, 1988), a consensus sequence (GXXXXGKS) for mononuclear binding protein in the nsP2 (Kaariainen *et al.*, 1987), a three amino acid motif (CWA) for the proteinase in the nsP2 (Hardy & Strauss, 1989), a net negative charge for nsP3 (Strauss *et al.*, 1984; Faragher *et al.*, 1988) and the RNA-dependent polymerase motif (GDD) near the C terminus of the nsP4 (Kamer & Argos, 1984). CHIK non-structural proteins, which are reported for the first time in this paper, possess all the features common to the alphaviruses, as shown in Results. In addition, putative cleavage sites for the polyproteins of alphaviruses were well conserved, except for the cleavage sites within E3 and E2 (data not shown).

The percentage identities of non-structural and structural polyproteins between CHIK and ONN were 85% compared with 42–70% between CHIK and other alphaviruses (Table 2). Thus, among the alphaviruses, ONN is the closest related virus to CHIK. Although RR and SF belong to the SF antigenic complex together with ONN and CHIK (Calisher *et al.*, 1980), RR and SF are more closely related to each other than to CHIK, as demonstrated by the percentage identities in Table 2. Phylogenetic analysis (Fig. 2) also demonstrated the same findings. These findings are in agreement with the data reported earlier by Lee *et al.* (1997) and Lanciotti *et al.* (1998), who analysed the nucleotide and amino acid sequences for the structural polyprotein of another CHIK strain (GenBank accession no. L37661). We observed that the amino acid sequence of their CHIK strain showed a 96% homology to our S27 strain (Table 2).

CHIK, RR and ONN possess very similar secondary structures at the 5′ NTR (Fig. 3), which may play a role in viral RNA replication. In contrast, we identified a significant difference between CHIK and ONN at the 3′ NTR, as demonstrated in Fig. 4. We found that CHIK, RR and BF contained similar RSEs in the 3′ NTR (Fig. 4A, B), whereas ONN possessed different types of RSE with different secondary structures, as previously reported by Levinson *et al.* (1990). One of the possible explanations for this difference is that the two viruses may have undergone divergent evolution from a common ancestral alphavirus. Also noteworthy is the

fact that CHIK, RR and BF use the *Aedes* mosquito as their vector, while ONN is the only known alphavirus to use the *Anopheles* mosquito. These RSEs may have a function in determining vector specificity during virus multiplication in the respective vector mosquitoes. This should be clarified in a separate study with genetically engineered recombinant viruses.

Some researchers have considered ONN as a subtype of CHIK based on a number of immunological studies revealing a one-way antigenic cross-reactivity between the two agents (Chanas *et al*., 1979; Calisher *et al*., 1980; Blackburn *et al*., 1995). However, a biological difference has also been recorded between the two viruses in that ONN does not replicate in an *Aedes aegypti* cell line (Chanas *et al*., 1979). Powers *et al*. (2000) reported that CHIK and ONN are two distinct viruses after carrying out phylogenetic analysis on the E1 protein and serological studies.

Comparing the nucleotide sequence of the E1 protein for previously published CHIK isolates RSU1, H15483, H2123 and Ag41855 (Powers *et al*., 2000) with CHIK in our study, the degree of identity ranged from 95 to 98%. The nucleotide sequence similarity of the E1 protein between the two isolates A234 and IbH12628 of ONN (Powers *et al*., 2000) was 97%. However, the nucleotide sequence identity of the E1 protein among CHIK isolates and ONN isolates ranged from 74 to 77%, and among CHIK isolates and VEE isolates P676 and Trinidad donkey (Kinney *et al*., 1992) ranged from 56 to 57%. Therefore, based on (i) nucleotide and amino acid homology among alphaviruses, (ii) data from phylogenetic analysis, and (iii) the characteristics of the RSEs found in the 3′ NTR, it can be concluded that CHIK and ONN, although closely related, are in fact two distinct viruses.

In this study, we found an internal poly(A) tract within the 3′ NTR of the CHIK genome. The length of the I-poly(A) varied from 19 to 106 adenine nucleotides among the six clones examined (Table 3). None of them was identical in length. This means that the internal poly(A) was generated not by an accidental insertion of an adenine oligonucleotide but, most likely, by an authentic polyadenylation capacity of CHIK RNA polymerase, which may generate poly(A) at the 3′ termini. We believe that in the region adjacent to the I-poly(A) of the S27 strain there must be a signal sequence that triggers or influences the occurrence of polyadenylation. We base this conclusion on the observation that in all six clones analysed, the poly(A) sequence started at the same position, i.e. nucleotide position 11438 (Fig. 1b).

Barr *et al*. (1997) reported that in the junction region of the vesicular stomatitis virus (VSV), the tetranucleotide 3′ AUAC 5′ followed by a U7 tract was implicated in the synthesis of poly(A) and termination of mRNA transcription. The process by which a poly(A) tail is templated by the U7 tract is called 'polymerase slippage'. In this process the VSV polymerase uses the U7 stretch as a template beginning a cycle that involves backward slippage of the polymerase on the nascent strand, followed by a cycle of nascent chain elongation and further slippage. It was also reported that AU-rich fragments preceding a U7 template, such as the tetranucleotides 3′ AUAA 5′, 3′ AUAU 5′ and 3′ AUAG 5′, can act as the signal for polymerase slippage but they are unable to signal termination (Barr & Wertz, 2001). In the 3′ NTR of CHIK, no consensus sequence was identified between the sequences adjacent to the I-poly(A) and 3′ poly(A), except that both sequences were AT(U) rich. We are currently unsure about the mechanism by which the I-poly(A) tract is created and whether the I-poly(A) is really generated by the same mechanism that synthesizes the 3′ poly(A). However, the existence of the I-poly(A) strongly suggests that the poly(A) is generated by a template-dependant mechanism similar to the 'polymerase slippage' seen for VSV, rather than simple nucleotidyl-terminal-transferase-type enzymic activity.

At the moment, it is unlikely that the I-poly(A) is a typical feature of the CHIK genome, because I-poly(A) does not exist in the 3′ NTR CHIK sequence obtained by Pfeffer *et al*. (1998), who reported a six-adenine sequence (U6 in the template RNA) at the same position of the same strain (S27). Also, two other CHIK strains isolated in Malaysia in 1998 possess four adenines at the same position (data not shown). Before this experiment, the CHIK strain S27 of our laboratory had been passaged more than 50 times using the *Aedes albopictus* clone C6/36 cell line (Igarashi, 1978). Because Pfeffer's CHIK (S27) and our CHIK possessed exactly the same nucleotide sequence for the 3′ NTR, except for the I-poly(A), it could be speculated that due to a history of multiple passage, one or more extra T (U in the template RNA) could have been added to the U6 tract of the viral RNA, creating the right template to initiate polyadenylation at that position. Further clarification of the process by which the I-poly(A) was generated may provide a clue to understanding fully the mechanism of polyadenylation among the alphaviruses.

The authors are grateful to Dr Shingo Inoue, Dr Basu Dev Pandey and Dr Mariko Tanaka for technical assistance. The first author is a recipient of the Monbusho scholarship from the Ministry of Education, Culture, Sports, Science and Technology of the Government of Japan for his stay and study in the Institute of Tropical Medicine, Nagasaki University. This study was supported by the Grant-in-Aid for Scientific Research (B) (No. 13576016).

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Barr, J. N. & Wertz, G. W. (2001). Polymerase slippage at vesicular stomatitis virus gene junctions to generate polyA is regulated by the upstream 3′-AUAC-5′ tetranucleotide : implications for the mechanism of transcription termination. *Journal of Virology* **75**, 6901–6913.

Barr, J. N., Whelan, S. P. J. & Wertz, G. W. (1997). *Cis*-acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *Journal of Virology* **71**, 8718–8725.

Blackburn, N. K., Besselaar, T. G. & Gibson, G. (1995). Antigenic relationship between Chikungunya virus strains and o'nyong-nyong virus using monoclonal antibodies. *Research Virology* **146**, 69–73.

Calisher, C. H., Shope, R. E., Brandt, W., Casals, J., Karabatsos, N., Murphy, F. A., Tesh, R. B. & Wiebe, M. E. (1980). Proposed antigenic classification of registered arboviruses. I. Togaviridae, Alphavirus. *Intervirology* **14**, 229–232.

Chanas, A. C., Hubalek, Z., Johnson, B. K. & Simpson, D. I. H. (1979). A comparative study of o'nyong-nyong virus with Chikungunya virus and plaque variants. *Archives of Virology* **59**, 231–238.

Diallo, M., Thonnon, J., Traore-Lamizana, M. & Fontenille, D. (1999). Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *American Journal of Tropical Medicine and Hygiene* **60**, 281–286.

Faragher, S. G., Meek, A. D. J., Rice, C. M. & Dalgarno, L. (1988). Genome sequences of a mouse-avirulent and a mouse-virulent strain of Ross River virus. *Virology* **163**, 509–526.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Felsenstein, J. (1993). PHYLIP: phylogeny inference package, version 3.5c. University of Washington, Seattle, Washington, USA.

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

Hardy, W. R. & Strauss, J. H. (1989). Processing the nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both *in cis* and *in trans*. *Journal of Virology* **63**, 4653–4664.

Igarashi, A. (1978). Isolation of a Singh's *Aedes albopictus* cell clone sensitive to dengue and Chikungunya viruses. *Journal of General Virology* **40**, 531–544.

Kaariainen, L., Takkinen, K., Keranen, S. & Soderlund, H. (1987). Replication of the genome of alphaviruses. *Journal of Cell Science* **7**(Suppl.), 231–250.

Kamer, G. & Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269–7282.

Kinney, R. M., Tsuchiya, K. R., Sneider, J. M. & Trent, D. W. (1992). Genetic evidence that epizootic Venezuelan equine encephalitis (VEE) viruses may have evolved from enzootic VEE subtype I-D virus. *Virology* **191**, 569–580.

Lanciotti, R. S., Ludwig, M. L., Rwaguma, E. B., Lutwama, J. J., Kram, T. M., Karabatsos, N., Cropp, B. C. & Miller, B. R. (1998). Emergence of epidemic o'nyong-nyong fever in Uganda after a 35-year absence: genetic characterization of the virus. *Virology* **252**, 258–268.

Lee, E., Stocks, C., Lobigs, P., Hislop, A., Straub, J., Marshall, I., Weir, R. & Dalgarno, L. (1997). Nucleotide sequence of the Barmah forest virus genome. *Virology* **227**, 509–514.

Levinson, R. S., Strauss, J. H. & Strauss, E. G. (1990). Complete sequence of the genomic RNA of o'nyong-nyong virus and its use in the construction of alphavirus phylogenetic trees. *Virology* **175**, 110–123.

Morita, K., Tanaka, M. & Igarashi, A. (1991). Rapid identification of dengue virus serotypes by using polymerase chain reaction. *Journal of Clinical Microbiology* **29**, 2107–2110.

Ou, J.-H., Strauss, E. G. & Strauss, J. H. (1981). Comparative studies of the 3′-terminal sequences of several alphavirus RNAs. *Virology* **109**, 281–289.

Ou, J.-H., Trent, D. W. & Strauss, J. H. (1982a). The 3′-noncoding regions of alphavirus RNAs contain repeating sequences. *Journal of Molecular Biology* **156**, 719–730.

Ou, J.-H., Rice, C. M., Dalgarno, L., Strauss, E. G. & Strauss, J. H. (1982b). Sequence studies of several alphavirus genomic RNAs in the region containing the start of the subgenomic RNA. *Proceedings of the National Academy of Sciences, USA* **79**, 5235–5239.

Ou, J.-H., Strauss, E. G. & Strauss, J. H. (1983). The 5′-terminal sequences of the genomic RNAs of several alphaviruses. *Journal of Molecular Biology* **168**, 1–15.

Page, R. D. M. (1996). Treeview: an application to display phylogenetic trees on personal computers. *Cabios* **12**, 357–358.

Pfeffer, M., Kinney, R. M. & Kaaden, O.-R. (1998). The alphavirus 3′-nontranslated region: size heterogeneity and arrangement of repeated sequence elements. *Virology* **240**, 100–108.

Porterfield, J. H. (1980). Antigenic characteristics and classification of Togaviridae. In *The Togaviruses*, pp. 13–46. Edited by R. W. Schlesinger. New York: Academic Press.

Powers, A. M., Brault, A. C., Tesh, R. B. & Weaver, S. C. (2000). Re-emergence of chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *Journal of General Virology* **81**, 471–479.

Ross, R. W. (1956). The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. *Journal of Hygiene* **54**, 177–191.

Sarkar, J. K., Chatterjee, S. N. & Chakravarty, S. K. (1964). Haemorrhagic fever in Calcutta: some epidemiological observations. *Indian Journal of Medical Research* **52**, 651–659.

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

Strauss, J. H. & Strauss, E. G. (1988). Evolution of RNA viruses. *Annual Review of Microbiology* **42**, 657–683.

Strauss, J. H. & Strauss, E. G. (1994). The alphaviruses: gene expression, replication and evolution. *Microbiological Reviews* **58**, 491–562.

Strauss, E. G., Rice, C. M. & Strauss, J. H. (1984). Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133**, 92–110.

Thaikruea, L., Charearnsook, O., Reanphumkarnkit, S., Dissomboon, P., Phonjan, R., Ratchbud, S., Kounsang, Y. & Buranapiyawong, D. (1997). Chikungunya in Thailand: a re-emerging disease? *Southeast Asian Journal of Tropical Medicine and Public Health* **28**, 359–364.

Van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R. & Wickner, R. B (editors) (2000). *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press.

Received 5 February 2002; Accepted 20 August 2002