Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Virology 410 (2011) 353–359

Contents lists available at ScienceDirect

Virology

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

Lineage replacement accompanying duplication and rapid fixation of an RNA element in the nsP3 gene in a species of alphavirus

John Aaskov *, Anita Jones, Wilson Choi, Kym Lowry, Emerald Stewart

Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, 4059, Australia

ARTICLE INFO

Article history:
Received 6 October 2010
Returned to author for revision
24 November 2010
Accepted 24 November 2010
Available online 23 December 2010

Keywords:
Alphavirus
Ross River virus
nsP3 gene
Lineage replacement

ABSTRACT

A sequence of thirty-six nucleotides in the nsP3 gene of Ross River virus (RRV), coding for the amino acid sequence HADTVSLDSTVS, was duplicated some time between 1969 and 1979 coinciding with the appearance of a new lineage of this virus and with a major outbreak of Epidemic Polyarthritis among residents of the Pacific Islands. This lineage of RRV continues to circulate throughout Australia and both earlier lineages, which lacked the duplicated element, now are extinct. Multiple copies of several other elements also were observed in this region of the nsP3 gene in all lineages of RRV. Multiple copies of one of these, coding for the amino acid sequence P*P*PR, were detected in the C-terminal region of the nsP3 protein of all alphaviruses except those of African origin. The fixation of duplications and insertions in 3′ region of nsP3 genes from all lineages of alphaviruses, suggests they provide some fitness advantage.

Introduction

Alphaviruses are positive sense RNA viruses that share a common ancestor with plant viruses in the tobamavirus, tobravirus and bromovirus families (Koonin and Dolja, 1993). New world alphaviruses commonly are associated with encephalitic disease in humans while infections with old world alphaviruses usually are associated with fever, rash and arthritis (Griffin, 2007). Following infection, the non-structural viral proteins (nsP1–4) of alphaviruses are translated directly from an open reading frame at the 5′ end of the viral genome while the structural proteins (C, E3, E2, 6K, E1) are derived from a 26S sub-genomic RNA produced by newly synthetised non-structural proteins (Strauss and Strauss, 1994). While the roles of non-structural proteins nsP1, 2 and 4 are well understood that of nsP3 is less clear. Furthermore, while alphavirus nsP1, 2 and 4 proteins share extensive sequence homology with proteins from other families of positive strand viruses, nsP3 does not (Ahlquist et al., 1985; Haseloff et al., 1984). nsP3 contains two conserved domains. The first (X or macro domain) is conserved among alphaviruses, coronaviruses, rubella and hepatitis E viruses (Koonin and Dolja, 1993) and the second is conserved among alphaviruses (Strauss and Strauss, 1994). nsP3 is highly phosphorylated, particularly the serine and threonine residues in the C-terminal region (Vihinen and Saarinen, 2000) and may act to attach the alphavirus replication complex (nsP1–4 proteins) to the cytoskeleton of the host cell (Frolova et al., 2006; Gorchakov et al., 2008). Semliki Forest viruses (SFV) can tolerate deletions of from 43 to 119 amino acids in the C-terminal region of their nsP3 proteins with only slight reductions in replication efficiency *in vitro* and in virulence for mice (Galbraith et al., 2006) and a 102 nucleotide deletion in this region of the nsP3 gene of Venezuelan encephalitis virus (VEEV) had no detectable effect on replication *in vitro* (Davis et al., 1989). Several members of the alphavirus family have an OPAL stop codon near the 3′ end of the nsP3 gene (Strauss et al., 1988) requiring read-through for production of the nsP4 polymerase. Duplicated amino acid elements have been observed in the C-terminal region of nsP3 of several alphavirus isolates (Meissner et al., 1999; Oberste et al., 1996; Strauss et al., 1988) but without any indication of when or where these events occurred and whether they were related to the epidemiology of the viruses concerned.

Ross River virus (RRV) employs complex, overlapping, urban and rural cycles of transmission involving multiple mosquito and vertebrate hosts but causes disease only in humans and horses (Russell, 2002). The nsP3 protein of a strain of Ross River virus (RRV) recovered from an Epidemic Polyarthritis patient in 2004 contained a duplication of the amino acid sequence HADTVSLDSTVS/L which had not been observed in any earlier isolates (Jones et al., 2010). The study described here was designed to determine whether the duplication of this element in this strain of RRV was an isolated event and, if not, when and where it had occurred and how quickly the change was fixed or removed.

Results and discussion

The amino acid sequence, HADTVSLDSTVS/L, which was duplicated in the nsP3 protein of RRV strain QML 1 recovered in 2004 (Jones et al., 2010), was duplicated in all examples of this lineage examined (lineage 3, Table 1, Fig. S1) but was present as only a single copy in the

* Corresponding author. Institute of Health and Biomedical Innovation, Queensland University of Technology, 60 Musk Avenue, Kelvin Grove, Brisbane, 4059, Australia.
E-mail address: j.aaskov@qut.edu.au (J. Aaskov).

0042-6822/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2010.11.025

Table 1
Amino acid repeat motifs in the nsP3 proteins of Ross River virus and their presence in the nsP3 proteins of other alphaviruses.

		Semliki Forest complex													WEE complex					
	Motif	Lineage I	Lineage 2	Lineage 3																
		RRV	RRV	RRV	RRV	RRV	RRV	RRV	RRV	GETV	SFV	MAYV	CHIKV	ONNV	BFV	SINV	AURV	WEEV	VEE	EEV
		T48	NB5092	F9073	MCLE	OREG	QML1	SNP51	PW14	AY702913	A7	AF237947	06–021	SC650	BH2193	SA.AR86	10315	71V-1658	OAX131	PE3.0803
		1959	1969	1979	1983	1989	2004	2009	2009											
332[a]	H	*H*[b]	*H*	H[c]	H	H	H	H	H	[d]										
	A	*A*	*A*	A	A	A	A	A	A											
	D	*D*	*D*	D	D	D	D	D	D											
	T	*T*	*T*	T	T	T	T	T	T											
	V	*V*	*V*	V	V	V	V	V	V											
	S	*S*	*S*	S	S	S	S	S	S											
	L	*L*	*L*	L	L	L	L	L	L											
	D	*D*	*D*	D	D	D	D	D	D											
	S	*S*	*S*	S	S	S	S	S	S											
	T	*T*	*T*	T	T	T	T	T	T											
	V	*V*	*V*	V	V	V	V	V	V											
	S	*S*	*S*	L/S	L/S	L/S	L/S	L/S	L/S											
383	P	*P*	*P*	*P*	*P*	*P*	*P*	P	P	*P*	*P*	*P*	*P*	*P*	*P*	*P*	P	P	P	P
	V	*V/I/V/T*	*V/1/V/T*	*V/I/V/T*	*V/I/V/T*	*V/I/V/T*	*V/I/M/T*	V/I/V/T	V/I/V/I	*I/V/V/A*	*I/V/T*	*V/I/V*	*V*	*I*	*I/V*	*V*	V	V/I/V	I/V	V/I/V
	P	*P*	*P*	*P*	*P*	*P*	*P*	P	P	*P*	*P*	*P*	*A*	*A*	*P*	*P*	P	P	P	P
	P	*P/A/A/T*	*P/T/A/T*	*P/A/A/T*	*P/A/A/T*	*P/A/A/T*	*P/A/A/T*	P/A/A/T	P/A/A/T	*P/A/A/R*	*P/A/A*	*P/A/A*	*P*	*P*	*A/A/P*	*P*	P	A/S/K	R/A/K	V/A/K
	P	*P*	*P*	*P*	*P*	*P*	*P*	P	P	*P/R/P/K*	*P*	*P*	*P*	*P*	*P*	*P*	P/L	P	P	P
	R	*R*	*R*	*R*	*R*	*R/H/R/R*	*R/H/R/R*	R	R/H/R/R	*R*	*R*	*R*	*R*	*R*	*R*	*R*	R	R	R	A/R
487	V	V	V	V	V	V	V	V	V	*V*										
	E	E	E	E	E	E	E	E	E	*E*										
	F	F/L	F/L	F/L	F/L	F/L	F/L	F/L	F/L	*L*										
	P	P	P	P	P	P	P	P	P	*P*										
	W	W	W	W	W	W	W	W	W	*W*										
	A	A/E	A/E	A/E	A/E	A/E	A/E	A/E	A/E	*E*										
	P	P	P	P	P	P	P	P	P	*P*										
	E	E	E	E	E	E	E	E	E	*E*										
	D	D	D	D	D	D	D	D	D	*D*										
	L	L/V	L/V	L/I	L/I	L/I	L/I	L/I	L/I	*L*										
521	D	D	D	D	D	D	D	D	D	D/G		*D*								
		−/K	−/K	−/K	−/K	−/K	−/K	−/K	−/K											
	I	I	I	I	I	I	I	I	I	I		*I*								
	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	T	*T*	T	T	T		T			
	F	F	F	F	F	F	F	F	F	F	F	*F*	F	F	F		F			
	G	G	G	G	G	G	G	G	G	G	G	*G*	G	G	G		G			
	D	D	D	D	D	D	D	D	D	D	D	*D*	D	D	D		D			

[a] Amino acid numbering from the N-terminal of RRV T48 nsP3.
[b] Single copy of the motif in italics.
[c] Multiple copies of motifs in bold type. Motif sequence from left to right from N-terminal to C-terminal of the nsP3 protein e.g HADTVSLDSTVL followed by HADTVSLDSTVS.
[d] Spaces indicate the motif was not observed in that virus.

two lineages of RRV which now are extinct (lineage 1 and 2, Table 1, Fig. S2). The C-terminal region of the nsP3 protein of RRV (amino acids 301–550) contained three additional elements that appeared to have been duplicated and one, P*P*PR, that appeared at four locations (Fig. 1A). Other elements contained fewer amino acids than the HADTVSLDSTVS one and the amino acid sequences were less conserved. Within the HADTVSLDSTVS/L element, there were three tri-peptides (TVS) which were not found elsewhere in the nsP3 protein of RRV suggesting they may have been the foot prints of previous duplication events in this region. While the sequence HADTVSLDSTVS was duplicated in all post-1979 strains of RRV studies, the other elements, that appeared at multiple sites, were observed in all lineages of RRV and in the nsP3 proteins of a number of other alphaviruses (Table 1).

The earliest example of a lineage 3 strain of RRV in which the element HADTVSLDSTVS/L was duplicated was recovered from an Epidemic Polyarthritis patient in Fiji in 1979 (Aaskov et al., 1981) at the beginning of an outbreak of infection that swept the Pacific region. The number of cases of RRV infection reported in Australia has climbed steadily from approximately 500 cases in 1980 to an average of approximately 5000 per year at present (Aaskov, 2009). Accompanying this increase in the number of cases in Australia has been the steady replacement of lineage 1 and 2 RRV by lineage 3 viruses (Jones et al., 2010). While there had been outbreaks of RRV infection in Australia prior to that in the Pacific, almost certainly caused by strains of RRV without this duplicated element in the nsP3 gene, these involved scores rather than tens of thousands of cases (Aaskov, 2009). However, we have been unable to identify a mechanism by which this change in the nsP3 gene could have conferred a significant fitness advantage on populations of RRV and it remains possible that one, or several, of the single nucleotide polymorphisms that distinguish the current lineage of RRV from the previous two (Jones et al., 2010) were responsible for these lineage replacements. There are precedents with other alphaviruses for epidemic potential to be determined by changes in only one or two nucleotides (Anischenko et al., 2006; Tsetsarkin et al., 2009). The task of evaluating the significance of the duplication of this element is made more difficult by the absence of RRV isolates from Epidemic Polyarthritis patients in Australia prior to 1983 (Aaskov et al., 1985) and the extensive passage of early lineages of RRV from pools of mosquitoes (which may have contained multiple infected insects) in the brains of suckling mice in order to recover isolates. Nonetheless, no changes to this element have been detected, and no further duplications in the nsP3 gene of RRV have been fixed, since 1979 (Table 1, Fig. S1).

A comparison of the nucleotide sequences of the nsP3 genes of the prototype strain of RRV (T48) and that in which the HADTVSLDSTVS repeat element was first observed (F9073) suggested three possible locations at which the duplication might have occurred i.e. 5′ to the original nucleotide sequence, 3′ to the original sequence or into the middle of it (Fig. 1B). Duplication of the sequence 5′ to its position in the RRV T48 genome would require changes to three nucleotides in the insert. Duplication of the sequence 3′ to its position in the T48 genome or by insertion into the middle of the original sequence would require nucleotide changes in both the T48 genome and in the duplicated element. If the insertion occurred 3′ to the ancestral sequence, the nucleotide sequences flanking the insertion site would have been almost identical (Fig. 1B). Duplication of this 36 nucleotide element converted a mildly disordered RNA structure in the RRV

A

```
  1  APSYRVRRTD ISGHAEEAVV NAANAKGTVG DGVCRAVARK WPDSFKGAAT PVGTAKLVRA
 61  NGMNVIHAVG PNFSTVTEAE GDRELAAAYR AVAGIINASN IKSVAIPLLS TGVFSGGKDR
121  VMQSLNHLFT AMDTTDADVV IYCRDKAWEK KIQEAIDRRT AVELVSEDIS LESDLIRVHP
181  DSCLVGRKGY SITDGKLHSY LEGTRFHQTA VDMAEISTLW PKLQDANEQI CLYALGESMD
241  SIRTKCPVED ADSSTPPKTV PCLCRYAMTA ERVARLRMNN XKAVIVCSSF PLPKYRIEGV
301  QKVKCDRVLI FDQTVPSLVS PRKYIPAAAS MHADTVSLDS TVLHADTVSL DSTVSTGSAW
361  SFPSEATYET MEVVAEVHHS EPPVPPPRRR RAQVTMHHQE LLEVSDMHTP IAARVEIPAY
421  DTAVVVERVA IPCTSEYATP IPAPRAARVV PVPAPRIQRA STYRVSPTPT PRVLRASVCS
481  VTTSAGVEFP WAPEDLEVLT EPVHCEMREP VELPWEPEDI DIQFGDFETP DKIQFGDIDF
541  DQF*LGRAGA
```

B

```
       961       971       981       991       1001      1011
                                          H  A  D  T  V  S  L  D  S
T48    CCAAGGAAGUACAUACCAGCCGCCGCCUCUACGCACGCAGAUACCGUGAGCUUGGAUUCU
T48    CCAAGGAAGUACAUACCAGCCGCCGCCUCUACGCACGCAGAUACCGUGAGCUUGGAU---
T48    CCAAGGAAGUACAUACCAGCCGCCGCCUCUACG--------------------------
F9073  CCAAGGAAAUACAUACCAGCCGCCGCCUCUAUGCACGCAGAUACCGUGAGUUUGGAUUCU

       1021      1031      1041      1051      1061      1071
       T  V  S  H  A  D  T  V  S  L  D  S  T  V  S
T48    ACAGUAUCC----------------------------------ACAGGAUCCGCGUGG
T48    ---------------------------------UCUACAGUAUCCACAGGAUCCGCGUGG
T48    ---------CACGCAGAUACCGUGAGCUUGGAUUCUACAGUAUCCACAGGAUCCGCGUGG
F9073  ACAGUAUUGCACGCAGAUACCGUGAGUUUGGAUUCUACAGUAUCUACAGGAUCCGCGUGG
             L
```

Fig. 1. Duplicated elements in the nsP3 protein/gene of Ross River virus strain F9073. (A) Duplicated amino acid sequences are represented in the same colour. Underlined sequences appear to be repeats within a repeat and are found nowhere else in nsP3. (B) Possible sites at which a 36 nucleotide element of the RRV T48 genome could have been inserted in the parental genome to give rise to the duplicated amino acid sequence. Amino acids coded by nucleotides of interest are shown above and below the nucleotide sequences. Codons which differ between RRV T48 (no repeat) and F9073 (duplicated element) are shown in pink. Similar nucleotide sequences flanking a putative insertion site are highlighted. Nucleotide numbering is from the 5′ end of the nsP3 gene.

genome into a more stable stem loop. (Fig. 2). Similar observations were made for RNA coding for a single and duplicated element in the 3′ region of the nsP3 gene from VEEV (Davis et al., 1989). However, given the additional energy required to unfold more stable RNA structures prior to translation or copying, it is difficult to imagine such changes would confer any fitness advantage. The element HADTVSLDSTVS/L differed from two of the others (PVPPPR and VEFPWAPEDL) which also appeared to have been duplicated in RRV in that it was not strongly hydrophobic. Even when variation occurred in the sequence of the two latter elements, the amino acid replacements usually were hydrophobic (Table 1). As these two elements were closer to the C-terminal of the nsP3 protein than the recently duplicated one, their hydrophobicity may indicate an association of this region of nsP3 with membranes or membrane-like structures in host cells (Gorchakov et al., 2008).

No inverted repeat nucleotide sequences were detected in the regions flanking the sites of the insertions, deletions or duplications in the nsP3 genes of alphaviruses and there were no A/U rich regions, which might be associated with polymerase slippage and recombination (Nagy and Simon, 1997), on either side of these changes either (Fig. S2). However, the sequences of the nucleotides on either side of one of the putative insertion site in RRV (Fig. 1) were almost identical as were the sequences flanking an insertion site in SFV (Fig. 3) but this was not the case in the other alphaviruses studied. The flanking nucleotide sequences in SFV were out of frame and so the similarities were not reflected in the amino acid sequence.

The changes observed in the nsP3 protein of RRV appeared less chaotic than those observed in this gene in other alphaviruses. Examples of duplicated elements, similar to those observed in RRV, but unique to particular families or lineages of alphaviruses are highlighted in Fig. 3. A full comparison of this region of the nsP3 protein of the major families of alphaviruses and the corresponding nucleotide sequences are shown elsewhere (Fig. S2) In both EEEV and VEEV, the duplicated element appeared 5′ to the original suggesting that the same may have occurred with the recently duplicated element in RRV nsP3. In contrast to RRV, the nsP3 genes of many other families of alphaviruses appeared to contain foreign genetic material. For example, there appeared to have been insertions of non-CHIKV RNA at two sites in the nsP3 gene of that virus. The amino acid element STITSLTHSQFDLSVDGE in CHIKV 06–021 was found in most strains of CHIKV but not in an example of one of the earliest lineages, ALSA 1. The amino acid sequence STITSLTH was identical to a region of a putative zinc finger protein from *Aedes aegypti* (Genbank XM001660684.1). The element GIADLAA in SFV (Y12518) was found nowhere else in the SFV polyprotein but appeared in a wide range of cellular proteins suggesting that host cell RNA could been inserted into this region of the SFV genome. Examples of what may represent foreign RNA inserted into the nsP3 genes of other alphaviruses have been reported previously (Davis et al., 1989, Oberste et al., 1996, Meissner et al., 1999) or are highlighted in EEV, SINV and VEEV in Fig. 3. In EEEV and SINV there appeared to be hotspots for insertion events with progressively larger elements being inserted at the same site of different lineages. As some repeats, e.g. P*P*PR, were observed in most lineages of alphaviruses (Powers et al., 2001), it is likely that the processes giving rise to them have been occurring for centuries. However, apart from two short ALAAR elements in an A-rich region, no repeat elements could be detected in the p150 gene/protein of rubella virus which has been suggested to be an antecedent of the alphavirus nsP3 gene (Koonin and Dolja, 1993).

The recent suggestion by Arrigo et al. (2010) that North American and South American lineages of EEV be reclassified as different species in the EEE complex is supported by an analysis of the amino acid sequences of the hypervariable region of their nsP3 proteins (Fig. 3). The EAEV/IH element is not duplicated in the North American lineage and this lineage appears to contain two, and possibly three, large insertions. Using similar criteria, there may be a case for making lineage 1E strains of VEEV a separate species in the VEE complex i.e. a large amino acid element is duplicated in VEEV lineages 1AB, 1C and

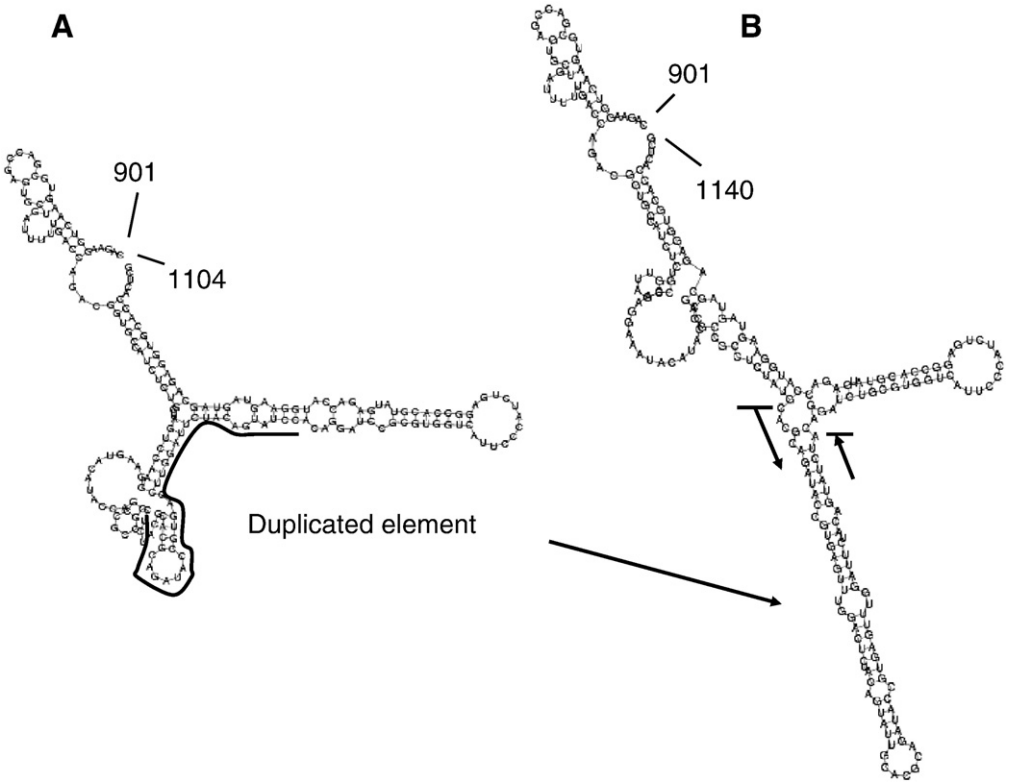

Fig. 2. Predicted secondary structure of the region of the RRV nsP3 gene in which a 36 nucleotide element was duplicated. (A) RRV T48 (B) RRV F9073 with the element duplicated. Nucleotide numbering refers to the position in the nsP3 gene of the respective viruses.

```
                    301       311       321       331       341       351
CHIKV 06-021        QKVKCSKVMLFDHNVPSRVSPREYRSSQESAQEASTITSLTHSQFDLSVDGEILPVPSDL
CHIKV ALSA-1        QKVKCSKVMLFDHNVPSRVSPREYRPSQESVQEA------------------ILPVPSDL

                    361       371       381       391       401       411
CHIKV 06-021        DADAPALEPALDDGATHTLPSTTGNLAAVSDWVMSTVPVAPPRRRRGRNLTVTCDEREGN
CHIKV ALSA-1        DADAPALEPALDDGAIHTLPSATGNLAAVSDWVMSTVPVAPPRRRRGRNLTVTCDER---

                    421       431       441       451       461       471
CHIKV 06-021        ITPMASVRFFRAELCPVVQETAETRDTAMSLQAPPSTATEPNHPPISFGASSETFPITFG
CHIKV ALSA-1        -----------AELCPVVQETAETRDTAMSLQAPPSTATELSHPPISFGAPSETFPITFG


                    361       371       381       391       401       411
EEE PE30803(IIIA)   S----PAVSMQSLGG----------SSTSEVIISEAEVH-----------DSDSDCSI
EEE PE17.0547(III)  S----PAVSMQSLGG----------SSTSEVIISEAEVH-----------DSDSDCSI
EEE PE240111(II)    S----PAVSMQSLGG----------SSTSDVVISEAEVH-----------DSDSECSI
EEE BEAR(IV)        S----PAISMQSLDGNTDTSVSGTALSSVASVTTIEAEIP-----------DSDSECSI
EEE NJ-60(I)        TSTNGSTTSIQSLGED-----QSASASSGAEISVDQVSLWSIPSATGFDVRTSSSLSLEQ

                    421       431       441       451       461       471
EEE PE30803(IIIA)   PPMP-FVVEAEVHASQGSQWSIPSASGFEIRE-SDDLG---SITRTPAISDHSVDLITFD
EEE PE17.0547(III)  PPMP-FVVEAEVHASQGSQWSIPSASGFEIREPLDDLG---SITRTPAISDHSADLITFD
EEEPE240111(II)     PPMP-FVVEAEVHASQGSHWSIPSASGFEIRELPEDRSISGSSTRASVISDHSVNLITFD
EEE BEAR(IV)        PPMP-FVVEAEVHASFGSQWSIPSATGFDIPEDCSVSSEGSISTHTSGVSGHSVNLITFD
EEE NJ-60(I)        PTFPTMVVEAEIHASQGSLWSIPSITGSETRVPSPPSQDSRPPTPSASASHTSVDLITFD

                    481       491       501       511       521       531
EEE PE30803(IIIA)   SVTDIFENFKQAPFQFLSDIRPIPAPRRRRE-PETDTQRFDKSEEKPVPKPRTRTAKYKK
EEE PE17.0547(III)  SVTDIFENFKQAPFQFLSDIRPIPAPRRRRE-PETDIQRFDKSEEKPVPKPRTRTAKYKK
EEE PE240111(II)    SVTDIFENFKQAPFQFLSEIRPIPAPRRRVGGLETDTKRYDKTEEKPIPKPRTRTTKYKQ
EEE BEAR(IV)        SVTDIFENFKQAPFQFLSDIRPIPAPRRHVVTPEDNQQRMR-----PIPKPRS-TVRYSK
EEE NJ-60(I)        SVAEILEDFSRSPFQFLSEIKPIPAPRTRVNNMSRSADTIK-----PIPKPRKSQVKYTQ


                    361       371       381       391       401       411
SFV(DQ189086)       QSCDIDSIYEPMAPIVVTADVHPEPAGIADLAADVHPEPADHVDLENPIPPPRPKRAAYL
SFV(Y12518)         QSCDIDSIYEPMAPIVVTADVHPEPA-------AVHPEPADHVDLENPIPPPRPKRAAYL

                    1141      1151      1161      1171      1181      1191
SFV(DQ189086)       GUACACCCUGAACCCGCAGGCAUCGCGGACCUGGCGGCAGAUGUGCAUCCUGAACCCGCA
SFV(Y12518)         GUGCACCCUGAACCCGCAGC--------------------UGUGCACCCUGAACCCGCA


                    301       311       321       331       341       351
SINVSAAR86          VQKVQCTKVVLFNPHTPAFVPARKYIEAPEQPAAPPAQAEEAPGVVATPTPPAA-DNTSL
OCKV                VQKVQCTKVVLFNPHTPAFVPARKYIEVPEQPAAPPAQDEEAPEAVATPAPPAA-DNTSL
SINVSW6562          VQKVQCTKVVLFNPQTPTFVPARKYIETPEQRITDVPTQEEPVNTAPEPTCTATGDNTSL

                    361       371       381       391       401       411
SINVSAAR86          DVTDISLDMEDSSEGSLFSSFSGSDN---------------YRRQVVVADVHAVQEPAPV
OCKV                DVTDISLDMDDSSEGSLFSSFSGSDNSITCMDRWSSGPSSLDRRQVVVADVHAVQEPAPI
SINVSW6562          DVTDISLDHEPSDQGSMSYDFAGSNSSIDSGMSWATPS---GRSVIVSAEVHAAQAPIPT

                    421       431       441       451       461       471
SINVSAAR86          PPPRLKKMARLAAA-RMQEEPTPPAS--TSSADESLHLSFDGVSISFGSLFDGEMARLAA
OCKV                PPPRLKKMARLAAASKTQEEPIPPAS--TSSADESLHLSFGGVSMSFGSLLDGEMARLAA
SINVSW6562          PPPRLKKLARLAAQAQLAAEETEPVTTDTTSEDESLHLSLNGMAMSFG------------

                    481       491       501       511       521       531
SINVSAAR86          AQPPASTCPTDVPMSFGSFSDGEIEELSRRVTESEPVLFGSFEPGEVNSIISSRSAVSFP
OCKV                AQPPA-TGPTDVPMSFGSFSDGEIEELSRRVTESEPVLFGSFEPGEVNSIISSRSAVSFP
SINVSW6562          -----------------SFTDGEVEELSRRKTNSEPVLFGSFEPGEVNSIISSRSAVSFP
```

Fig. 3. Variation in the amino acid sequences of nsP3 proteins of different lineages within families of alphaviruses. Repeated elements are shown in bold type and what appear to be inserts of foreign sequence are shaded in grey.

```
                  301       311       321       331       341       351
VEEV71-180(1AB)  RITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVDETPEPSAENQSTEGTPEQPPLIT
VEEV8131(1D)     RITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPTVEETPEPPAENQPTEGTPEQPTLIT
VEEVPMCHo5(1C)   RITGVQKIQCSQPILFSPKVPAYIHPRKYLVETPPVEETPESPAENQSTEGTPEQPALVN
VEEVOAX131(1E)   RITGVQKIQCSHPILFSPKVPEYIHPRKYLADA--------ASANNEAAESTS-----VD

                  361       371       381       391       401       411
VEEV71-180(1AB)  EDETRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHG-PPSVSSSSWSIPHASDF
VEEV8131(1D)     VDETRTRTPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHG-PPSASSSSWSIPHASDF
VEEVPMCHo5(1C)   VDATRTRMPEPIIIEEEEEDSISLLSDGPTHQVLQVEADIHG-SPSVSSSSWSIPHASDF
VEEVOAX131(1E)   VQPQLEESPENTEQLVEEEDSISVLSEDTPHQEHQVEAEVHRFSASAQSSSWSIPRASDF

                  421       431       441       451       461       471
VEEV71-180(1AB)  DVDSLSILDTLEGASVTSGATSAETNSYFAKSMEFLARPVPAPRTVFR------------
VEEV8131(1D)     DVDSLSILDTLEGASVTSEEASVETNSYFARSMEFLARPVPAPRTVFR------------
VEEVPMCHo5(1C)   DVDSLSILDTLDGASVTSEAASAETNSYFARSMEFRARPVPAPRTVFR------------
VEEVOAX131(1E)   DVESLSVLESL-GANDTISMESSSNETALALRTIFRTPPIPRPRVQSTSTDVDSISALES

                  481       491       501       511       521       531
VEEV71-180(1AB)  -NPPHPAPRTRTPSLAPSRAC**SRTSLVSTPPGVNRVITREELEA**LTPSRTP-----SRSV
VEEV8131(1D)     -NPPQPAPRTRTPSLAPSRAS**SRISLVSNPPGVNRVITREELEA**LTPSRTP-----SRSV
VEEVPMCHo5(1C)   -NPPHPAPRTRTPPLAHSRAS**SRTSLVSTPPGVNRVITREELEA**LTPSRAP-----SRSA
VEEVOAX131(1E)   CDSTSDARSVDSDETDVSIFDKRLEFRARPVPAPRTKFRTPPVPKPRARRPFHPLSSRSS

                  541       551       561       571
VEEV71-180(1AB)  **SRTSLVSNPPGVNRVITREEFEA**FVAQQQRFDAGA
VEEV8131(1D)     **SRTSLVSNPPGVNRVITREEFEA**FVAQQQRFDAGA
VEEVPMCHo5(1C)   **SRTSLVSNPPGVNRVITREEFEA**FVAQQQRFDAGA
VEEVOAX131(1E)   **SRSSLASNPPGVNRVITREEFEA**FVAQQQRFDAGA
```

Fig. 3 (*continued*).

1D but not in 1E and lineage 1E viruses contain three large sequences not found in the other lineages of VEEV.

The changes observed in the C-terminal region of the nsP3 gene/protein of RRV and other alphaviruses bore some similarities to those in defective interfering (D.I.) particles of SINV and SFV i.e. linear repeats and the insertion of foreign nucleotide sequences (Lehtovaara et al., 1981; Tsiang et al., 1985) raising the possibility that the processes giving rise to the hypervariability in nsP3 genes are similar to those that give rise to alphavirus DI particles. These observations and earlier studies (Davis et al., 1989, Lehtovaara et al., 1981; Tsiang et al., 1985) suggest

Table 2
Alphaviruses analysed in this study.

Virus	Strain(lineage)	Year of isolation	Source	Location	Accession number	Amino acids in nsP3
AURAV	BeAR 10315	1959	*Culex sp.*	Brazil	AF126284	544
BFV	BH2193	1974	*Culex sp.*	Australia	U73745.1	470
CHIKV	06–021	2006	Human	Reunion	AM258992	530
	ALSA-1	1986		India	HM045806.1	495
EEEV	NJ-60 (I)	1959	*Culiseta sp.*	USA	EF568607	559
	PE24.0111 (II)	2000	Mosq.	Peru	DQ280401	539
	PE17.0547 (III)	1998	Mosq.	Peru	DQ280397	536
	PE3.0803 (IIIA)	1996	Mosq.	Peru	DQ280386	535
	BeAR436087 (IV)	1985	*Culex sp.*	Brazil	EF151503	545
GETV			Porcine	Korea	AY702913	524
GETV	Sagiyama M6/Mag32	1956	*Culex sp.*	Japan	AB032553	524
MAYV					AF237947.1	492
ONNV	SG650	1996	Human	Uganda	AF079456	569
RRV	T48	1959	*Aedes sp.*	Australia	GQ433359	538
	NB5092	1969	*Aedes sp.*	Australia	M20162	538
	F9073	1979	Human	Fiji		550
	MCLE	1983	Human	Australia		550
	OREG	1989	Human	Australia		550
	QML 1	2004	Human	Australia	GQ433354	550
	SNP 51	2009	Human	Australia		550
	PW 14	2009	Human	Australia		550
SFV	A7				Y12518.1	475
	SK	1970		Finland	DQ189086	482
SINV	S.A.AR86			South Africa	U38305	543
	SW6562		Mosq.	Australia	AF429428	523
	Ockelbo Edsbyn				M69205.1	558
VEEV	71–180 (1AB)	1971	Equine	USA	AF069903.1	557
	PMCHo5 (1C)				U55345.2	557
	8131 (1D)		Human	Peru	DQ390224.2	557
	OAX131 (1E)				AF448536.1	562
WEEV	71V-1658	1971	USA	Equine	AF214040.1	532
	AG80–646	1980	*Culex sp.*	Argentina	GQ287646.1	529

that the hypervariability of the nsP3 gene and the generation of alphavirus DI particles both could be due to recombination as a result of RNA template switching by nsP4. The duplication events in EEEV, VEEV and possibly RRV occurred 5′ to the original element, suggesting that recombination could have occurred during synthesis of negative strand RNA. Perhaps nsP4 is more prone to template switching when it is associated with the uncleaved nsP1–3 polyprotein to synthetise negative strand RNA than when it is complexed with nsP1, nsP2 and nsP3 proteins to produce positive strand RNA.

The association of changes in the envelope proteins of alphaviruses with outbreaks of disease (Anischenko et al., 2006; Tsetsarkin et al., 2009) has focussed attention on the structural proteins of this family of viruses. However, while changes to structural proteins have the potential to influence the entry into, and the egress from, infected cells by virions, changes to non-structural proteins have the potential to have profound effects on the amount of virus produced and on the fitness of those virions e.g. depending on the fidelity of the replication of viral genomes (Pfeiffer and Kirkegaard, 2005). The observation that all alphaviruses appear to insert pieces of autologous and or heterologous RNA into the 3′ region of their nsP3 genes and that some of these changes spread rapidly throughout lineages of these viruses suggests that there is some evolutionary benefit accruing from this process. What this benefit might be remains to be elucidated.

Materials and methods

Viruses

Strains of RRV (Table 2) were obtained from the collection at the World Health Organisation Collaborating Centre for Arbovirus Reference and Research at the Queensland University of Technology. Nucleotide sequences for other alphaviruses were obtained from Genbank.

Nucleotide sequencing and analysis

RNA was extracted from RRV in the supernatant of cultures of infected Vero cells with QIAamp viral RNA minicolumns (Qiagen), according to the manufacturer's instructions. The RNA was reverse transcribed with Superscript III reverse transcriptase (Invitrogen) and random hexanucleotide primers (Boehringer). The resultant cDNA was amplified using a mixture of *Taq* and *Pwo* polymerases (Expand Long Template DNA polymerase; Roche) and RRV nsP3 specific primers (Table 3). The PCR product was analysed in 1.5% w/v agarose–Tris-acetate–EDTA gels, and bands of cDNA of interest were recovered and purified with QIAquick gel extraction kits (Qiagen) according to the manufacturer's instructions. The cDNA was sequenced at the Australian Genome Research Facility (Brisbane) using di-deoxy dye termination technology (Applied Biosystems). Sequences were aligned and analysed with software (Clustal W, DNAdist, Seqboot, Consense, Neighbour, M-Fold) available from the Australian National Genome Information Service (http://biomanager.info/). The one letter amino acid code has been used to identify amino acids.

Supplementary materials related to this article can be found online at doi: 10.1016/j.virol.2010.11.025.

Table 3
Oligonucleotide primers used to amplify and sequence the nsP3 gene of Ross River virus.

P3537[a,b]	CAGGGCGAGAGGGTAGAATGG	3534–3554[c]
cP4200[c]	CATTTTCTCGCCACCGCTCTG	4175–4195
cP4486	GCGTCCGTGGTGTCCATTGC	4460–4479
P4	TCACTTGAGTCTGATTTGATACGGG	4581–4605
P4774	GCATTGGGTGAGAGTATGGACAG	4773–4795
cP3	ATTTGCTTCTGATACTGTCCATACTCTC	4782–4809
P4854	GTTCCGTGTCTGTGTAGGTATGC	4851–4873
P4932	GTGTGCTCTTCATTCCCTTTACC	4931–4953
P5307	GCTGTTGTAGCGGAGAGAGTGG	5304–5325
cP6097	CCTCTGTCGGGTAATTGGCTTC	6075–6096

[a] P — sense primer.
[b] cP — complimentary primer.
[c] Numbering refers to that for nucleotides in RRV T48.

Acknowledgements

This study was supported by grants from the Cook Estate and the National Health and Medical Research Council of Australia.

References

Aaskov, J.G., 2009. Ross River virus: epidemic polyarthritis. In: Barrett, A.D.T., Stanberry, L.R. (Eds.), Vaccines for Biodefence and Emerging and Neglected Diseases. Elsevier, Amsterdam, pp. 632–645.
Aaskov, J.G., Mataika, J.U., Lawrence, G.W., Rabukawaqa, V., Tucker, M.M., Miles, J.A., Dalglish, D.A., 1981. An epidemic of Ross River virus in Fiji, 1979. Am. J. Trop. Med. Hyg. 30, 1053–1059.
Aaskov, J.G., Ross, P.V., Harper, J.J., Donaldson, M.D., 1985. Isolation of Ross River virus from epidemic polyarthritis patients in Australia. Aust. J. Biol. Med. Sci. 63, 587–597.
Ahlquist, P., Strauss, E.G., Rice, C.M., Strauss, J.H., Haseloff, J., Zimmern, D., 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to non-structural proteins from several RNA plant viruses. J. Virol. 53, 536–542.
Anischenko, M., Bowen, R.A., Paessler, S., Austgen, L., Greene, I.P., Weaver, S.C., 2006. Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. Proc. Natl Acad. Sci. U. S. A. 103, 4994–4999.
Arrigo, N.C., Adams, A.P., Weaver, S.C., 2010. Evolutionary patterns of Eastern Equine Encephalitis virus in North versus South America suggest ecological differences and taxonomic revision. J. Virol. 84, 1014–1025.
Davis, N.L., Willis, L.V., Smith, J.F., Johnston, R.E., 1989. In vitro synthesis of infectious Venezuelan Equine Encephalitis virus RNA from a cDNA clone: analysis of a viable deletion mutant. Virology 171, 189–204.
Frolova, E., Gorchakov, R., Garmashova, N., Atasheva, S., Vergara, L.A., Frolov, I., 2006. Formation of nsP3 specific protein complexes during Sindbis virus replication. J. Virol. 80, 4122–4134.
Galbraith, S.E., Sheahan, B.J., Atkins, G.J., 2006. Deletions in the hypervariable domain of the nsP3 gene attenuates Semliki Forest virus virulence. J. Gen. Virol. 87, 937–947.
Gorchakov, R., Garmashova, N., Frolova, E., Frolov, I., 2008. Different types of nsP3-containing protein complexes in Sindbis virus-infected cells. J. Virol. 82, 10088–10101.
Griffin, D.E., 2007. Alphaviruses. In: Knipe, D.M., Howley, P.M. (Eds.), Lippincott, Philadelphia, pp. 1023–1067.
Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R., Kaesberg, P., 1984. Striking similarities in amino acid sequence among non-structural proteins encoded by RNA viruses that have dissimilar genomic organisation. Proc. Natl. Acad. Sci U. S. A. 81, 4358–4362.
Jones, A., Lowry, K., Aaskov, J., Holmes, E.C., Kitchen, A., 2010. Molecular evolutionary dynamics of Ross River virus and implications for vaccine efficacy. J. Gen. Virol. 91, 182–188.
Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28, 375–430.
Lehtovaara, P., Soderlund, H., Keranen, S., Pettersson, R.F., Kaariainen, L., 1981. 18S defective interfering RNA of Semliki Forest virus contains a triplicated linear repeat. Proc. Natl Acad. Sci. USA 78, 5353–5357.
Meissner, J.D., Huang, C.Y., Pfeffer, M., Kinney, R.M., 1999. Sequencing of protype viruses in the Venezuelan equine encephalitis antigenic complex. Virus Res. 64, 43–59.
Nagy, P.D., Simon, A.E., 1997. New insights into the mechanism of RNA recombination. Virology 235, 1–9.
Oberste, M.S., Parker, M.D., Smith, J.F., 1996. Complete sequence of Venezuelan equine encephalitis virus subtype 1E reveals conserved and hypervariable domains within the C terminus of nsP3. Virology 219, 314–320.
Pfeiffer, J.K., Kirkegaard, K., 2005. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. 1, e11.
Powers, A.M., Brault, A.C., Shirako, Y., Strauss, E.G., Kang, W., Strauss, J.H., Weaver, S.C., 2001. Evolutionary relationships and systematics of the alphaviruses. J. Virol. 75, 10118–10131.
Russell, R.C., 2002. Ross River virus: ecology and distribution. Annu. Rev. Entomol. 47, 1–31.
Strauss, E.G., Levinson, R., Rice, C.M., Dalrymple, J., Strauss, J.H., 1988. Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: sequence and comparison with those of oyther alphaviruses. Virology 164, 265–274.
Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: gene expression, replication and evolution. Microbiol. Rev. 58, 491–562.
Tsetsarkin, K.A., McGee, C.E., Volk, S.M., Vanlandingham, D.L., Weaver, S.C., Higgs, S., 2009. Epistatic roles of E2 glycoprotein mutations in adaption of Chikungunya virus to *Aedes albopictus* and *Ae.* Aegypti mosquitoes. PLoS ONE 4, e6835.
Tsiang, M., Monroe, S.S., Schlesinger, S., 1985. Studies of defective interfering RNAs of Sindbis virus with and without $tRNA^{Asp}$ sequences at their 5′ termini. J. Virol. 54, 38–44.
Vihinen, H., Saarinen, J., 2000. Phosphorylation site analysis of Semliki Forest virus non-structural protein 3. J. Biol. Chem. 275, 27775–27783.