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.