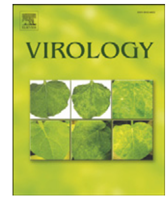
CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

Virology 399 (2010) 98–108

Contents lists available at ScienceDirect

Virology

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



Ngaingan virus, a macropod-associated rhabdovirus, contains a second glycoprotein gene and seven novel open reading frames

Aneta Gubala [a,b,c,d,*], Steven Davis [e], Richard Weir [e,d], Lorna Melville [e,d], Chris Cowled [a,d], Peter Walker [a], David Boyle [a,d]

[a] CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Victoria, Australia
[b] School of Chemistry and Molecular Sciences, University of Queensland, St. Lucia, Queensland, Australia
[c] Human Protection and Performance Division, Defence Science and Technology Organisation, Melbourne, Australia
[d] Australian Biosecurity Cooperative Research Centre for Emerging Infectious Disease, Brisbane, Queensland, Australia
[e] Northern Territory Department of Regional Development, Primary Industry, Fisheries and Resources, Berrimah Veterinary Laboratories, Berrimah, Northern Territory, Australia

A R T I C L E I N F O

Article history:
Received 16 October 2009
Returned to author for revision
16 November 2009
Accepted 14 December 2009
Available online 20 January 2010

Keywords:
Rhabdovirus
Ngaingan
Rhabdoviridae
Arbovirus
Evolution
Genome
Characterization
(-) ssRNA
Virus
Serology

A B S T R A C T

Ngaingan virus (NGAV) was isolated from a pool of biting midges that were collected in the tropics of northern Australia. Reported here is the full-length sequence of the NGAV genome, which, at over 15.7 kb, is the largest in any rhabdovirus described to date and contains 13 genes, the highest number of genes observed in any (−) ssRNA virus. Seven of these putative genes show no significant homology to known proteins. Like viruses in the genus *Ephemerovirus*, NGAV possesses a second glycoprotein gene (G_{NS}). Phylogenetic analyses, however, place NGAV within the yet to be classified "Hart Park" group containing Wongabel and Flanders viruses, which do not contain a second glycoprotein gene. Screening of various animal sera from northern Australia has indicated that NGAV is currently circulating in macropods (wallabies, wallaroos and kangaroos), highlighting the need for further studies to determine its potential to cause disease in these species.

Crown Copyright © 2009 Published by Elsevier Inc. All rights reserved.

Introduction

Ngaingan virus (NGAV) was isolated in 1970 from a pool of biting midges that were collected for surveillance for arboviruses at the low-lying plains of the Mitchell River Aboriginal community (Kowanyama; lat/long −15° 28′ S, 141° 44′ E), Gulf of Carpentaria, northern Queensland (Doherty et al., 1973). At the time, the midge pool was thought to consist of only *Culicoides brevitarsis* (which do not favour feeding on macropods), but later studies of *Culicoides* spp. from the region suggested that *C. actoni* (which have a wide feeding range including macropods and cattle) were probably also present (Kay et al., 1978). Another yet uncharacterised rhabdovirus, Almpiwar virus, was also isolated from this area a number of times from skinks. Morphological examination of NGAV revealed bullet-shaped virions typical of the *Rhabdoviridae* (Karabatsos, 1985). Experimental infection of mosquitoes showed that it can multiply in *Aedes aegypti*, a characteristic common of arboviruses (Carley et al., 1973). Initial characterisation suggested that NGAV was related antigenically to Tibrogargan virus, another uncharacterised Australian rhabdovirus (Calisher et al., 1989). However, recent phylogenetic analysis of a short L gene fragment indicated that it is a member of the Hart Park group, which also includes Wongabel (WONV) and Parry Creek viruses, both from Australia, and Flanders virus (FLAV) isolated from mosquitoes and birds in the USA (Bourhy et al., 2005).

Early serologic surveys suggested that NGAV infects wallabies, kangaroos and possibly cattle (Doherty et al., 1973). Whilst a recent outbreak of blindness in kangaroos and wallabies has been associated with the Wallal and Warrego orbiviruses (Hooper et al., 1999; Reddacliff et al., 1999), viral diseases in these species are poorly studied and it remains to be determined whether NGAV also causes disease.

The genomes of rhabdoviruses always contain genes encoding the five structural proteins, nucleoprotein (N), phosphoprotein (P), matrix protein (M), transmembrane glycoprotein (G) and RNA-dependent RNA polymerase (L). The complete genome sequences

* Corresponding author. Human Protection and Performance Division, Defence Science and Technology Organisation, PO Box 4331, Melbourne 3001, Victoria, Australia. Fax: +61 3 9626 8410.
E-mail address: ania.gubala@dsto.defence.gov.au (A. Gubala).

0042-6822/$ – see front matter. Crown Copyright © 2009 Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2009.12.013

obtained for a number of rhabdoviruses demonstrate that some possess numerous additional genes or ORFs that putatively encode small proteins with unknown functions (Basurco and Benmansour, 1995; Dietzgen et al., 2006; Gubala et al., 2008; Huang et al., 2003; McWilliam et al., 1997; Reed et al., 2005; Revill et al., 2005; Scholthof et al., 1994; Springfeld et al., 2005; Tanno et al., 2000; Walker et al., 1991; Walker et al., 1992; Wang et al., 1994; Wang and Walker, 1993). In addition, the two characterised ephemeroviruses, bovine ephemeral fever virus (BEFV) and Adelaide River virus (ARV), are known to contain a second glycoprotein (G_{NS}) with an unknown function (Walker et al., 1991; Walker et al., 1992; Wang and Walker, 1993). Due to the large number of evolutionarily divergent rhabdoviruses for which little or no genome sequence information is available, many remain unassigned to any of the six genera (*Lyssavirus*, *Vesiculovirus*, *Ephemerovirus*, *Novirhabdovirus*, *Cytorhabdovirus* and *Nucleorhabdovirus*) recognised currently.

This paper describes the sequence of the complete genome of NGAV, which at 15,764 nt is currently the largest known within the *Rhabdoviridae*. The NGAV genome possesses 13 discrete genes predicted to encode proteins, which is the most of any known (−) ssRNA virus and highlights the levels of complexity rhabdoviruses are capable of reaching. In addition to the five genes coding for structural proteins typical of rhabdoviruses, NGAV possesses a gene encoding a second glycoprotein, which has previously only been observed in members of the genus *Ephemerovirus*, and seven encoding novel proteins. Phylogenetic analyses support assignment of a seventh genus within the *Rhabdoviridae* comprising of NGAV, WONV and FLAV. Serologic data indicate that NGAV is still circulating in some animal species in northern Australia.

Results

Complete genomic sequence of NGAV

The complete genome of NGAV is 15,764 nt in length and contains the five typical rhabdovirus genes N, P, M, G and L (Fig. 1A). In addition, it contains seven relatively short ORFs predicted to encode novel proteins with no significant similarity at the amino acid level to any proteins currently in GenBank, as well as a longer ORF consistent in genome position and sequence to a second glycoprotein gene, a feature previously only observed in the ephemeroviruses BEFV and ARV (Walker et al., 1991; Walker et al., 1992; Wang and Walker, 1993). Each of the ORFs is bounded by recognisable transcription control sequences (with the exception of U1/U2 and U5/U6 which appear to be bicistronic, as described later), suggesting that all are transcribed. Details of the genes, ORFs, putative proteins and untranslated regions (UTRs) are collated in Table 1.

Although a consensus sequence of high confidence was generated for the whole genome, four nucleotide positions remained ambiguous despite repeated sequencing. Conflicts at positions 5842, 8687 and 12,853 within the G, U6 and L genes, respectively, were nonsynonymous and resulted in amino acid transitions, and the conflict at position 3153 introduced a stop codon within the U3 gene. These conflicts are noted in the NGAV GenBank entry (accession no. FJ715959).

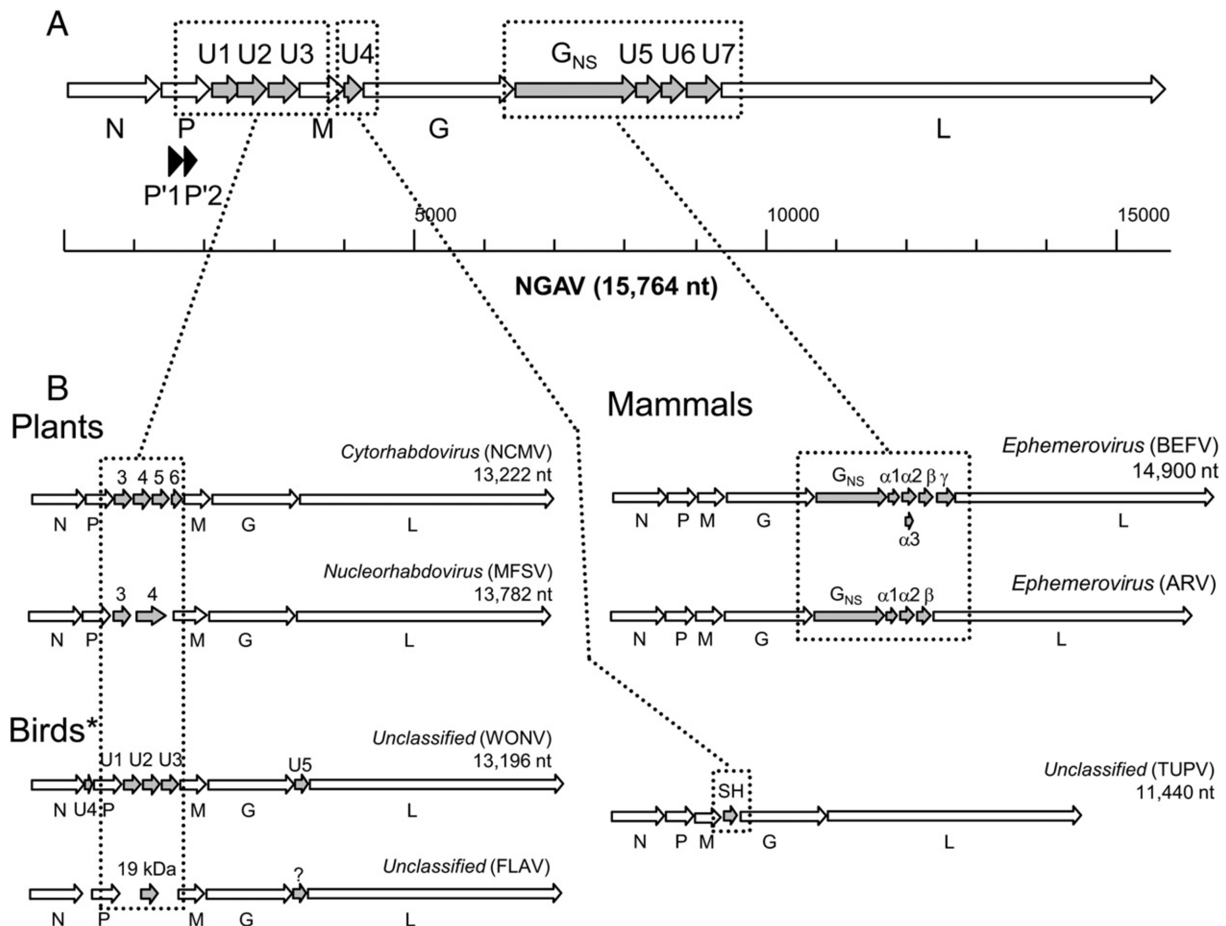


Fig. 1. Comparison of NGAV genome organisation with other rhabdoviruses. Common rhabdovirus proteins are depicted by hollow arrows and novel ORFs by shaded arrows. (A) Organisation of the NGAV genome. ORFs P'1 and P'2 (black arrow heads) overlap with P in different frames. (B) Genomic locations of novel NGAV ORFs compared with other rhabdoviruses. *Serologic evidence suggests that WONV infects birds although it has never been isolated from this host. FLAV has been isolated from birds and mosquitoes.

Table 1
Predicted genes and putative proteins of NGAV.

Protein	Gene length (nt)	ORF length (nt)	5′ UTR (nt)	3′ UTR (nt)	Protein length (aa)	Protein calculated Mr (kDa)	Isoelectric point (pI)
N	1331	1290	26	15	429	48.2	6.4
P	727	696	10	21	231	26.1	5.9
P'1	–	207	–	–	68	7.7	11.2
P'2	–	177	–	–	58	6.9	11.4
U1	781*	420	12	–	139	15.8	6.9
U2	781*	402	–	6	133	15.2	7.9
U3	463	441	11	11	146	17.1	4.9
M	639	615	13	11	204	23.4	8.0
U4	264	246	11	7	81	9.7	5.0
G	2161	2142	13	6	713	80.9	6.5
G_{NS}	1727	1707	13	7	568	65.0	6.7
U5	716*	345	10	-	114	12.6	8.4
U6	716*	339	-	11	112	13.0	10.0
U7	486	462	13	11	153	18.3	10.2
L	6374	6318	11	45	2105	242	7.2

Gene positions and features are presented in the antigenome direction. 3′ UTRs include polyadenylation sequences. *U1 and U2 are encoded by a bicistronic mRNA, as are U5 and U6.

N protein

The NGAV N protein (429 aa) is predicted to contain numerous phosphorylation sites and thus, like most rhabdovirus N proteins, is probably highly phosphorylated. Four *N*-myristoylation sites and an amidation site are also predicted as well as a C-terminal microbody targeting signal (SHA motif), which is not typically found in viruses. Proteins that contain this signal are characteristically directed to microbodies, which are small organelles that specialize in cell metabolism (de Hoop and Ab, 1992). Examination of all rhabdovirus N protein sequences in GenBank indicated that only the unclassified insect virus SIGMAV contains a microbody targeting signal (TRV motif). The presence of this signal could indicate a novel and yet undescribed function of rhabdoviral N proteins potentially involved in altering host-cell metabolism. The NGAV N protein also possesses a sequence (*GLSNRSPYSA*) highly conserved with the G(L/I)SXKSPYSS RNA binding motif present in the cognate protein of ephemeroviruses, vesiculoviruses and lyssaviruses (Crysler et al., 1990; Kouznetzoff et al., 1998; Tordo et al., 1986; Walker et al., 1994).

P protein

The 231-aa P protein of NGAV is predicted to contain nine phosphorylation sites. Whilst x-ray crystallography and NMR studies suggest an overall common modular organisation of rhabdovirus P proteins (Mavrakis et al., 2004; Ribeiro et al., 2008), direct comparisons are difficult as they share little sequence similarity. In VSIV, a nonphosphorylated 162-aa stretch located centrally in the N protein is believed to play an important role in virion assembly (Das and Pattnaik, 2005) but based on the predicted phosphorylation pattern, no recognisable cognate domain exists in the NGAV P protein. The NGAV P gene contains two additional ORFs that overlap P in a similar way as the additional ORF detected in several other rhabdoviruses referred to as either C or P' (Kretzschmar et al., 1996; Peluso et al., 1996; Spiropoulou and Nichol, 1993; Springfeld et al., 2005) (Fig. 1A). Examination of the protein products of these ORFs in vesiculoviruses and paramyxoviruses suggests that they are nonessential but can affect the host immune system by suppressing apoptosis, assist virus budding and regulate transcription and replication (Irie et al., 2008; Kretzschmar et al., 1996; Nagai and Kato, 2004; Peluso et al., 1996). As the NGAV P protein contains two putative overlapping ORFs, these have been named P'1 and P'2. In size and pI, both the P'1 and P'2 protein products (Table 1) are consistent with other rhabdovirus C and P' proteins. Both ORFs also possess alternate downstream start codons with potential to produce shorter isoforms similarly to VSNJV and VSIV (Kretzschmar et al., 1996; Spiropoulou and Nichol, 1993).

M protein

The 204-aa M protein of NGAV is predicted to contain several phosphorylation sites and an *N*-myristoylation site. A common feature of matrix proteins in various enveloped viruses is the presence of proline-rich L-domains near the amino terminus (PPXY, PT/SAP, YXXL and FPIV) (Chen and Lamb, 2008; Harty et al., 1999). The domains interact with host cell proteins and also help mediate virus budding and have been studied extensively in RABV, VSNJV and VSIV (Harty et al., 1999; Wirblich et al., 2008). Although not apparent in the NGAV M protein, the corresponding M protein region (residues 30 to 54) is also proline-rich, suggesting similar functionality masked by low amino acid sequence identity.

G protein

The 713-aa G protein of NGAV is the largest of any characterised rhabdovirus, some 44 aa larger than the G protein (669 aa) of the plant-infecting rice yellow stunt rhabdovirus (RYSV) (Luo and Fang, 1998), and 53 aa to 95 aa larger than the G proteins of BEFV, ARV, FLAV and WONV. Most additional NGAV G protein sequence occurs in its amino-terminal (ectodomain) region.

The NGAV G protein is predicted to contain characteristics typical of the class I transmembrane glycoproteins of other rhabdoviruses including a signal peptide (residues 1 to 19), ectodomain (residues 20 to 636), 18-aa transmembrane region (residues 637 to 654) and a cytoplasmic tail (residues 655 to 713). Rhabdovirus G proteins typically contain two to six *N*-glycosylation sites, although numerous newly sequenced viruses are predicted to contain more (Coll, 1995). The NGAV G protein is predicted to contain four *N*-glycosylation sites, three of which are located in the ectodomain. As 30 putative phosphorylation sites exist throughout the protein, it is also likely to be phosphorylated as in other rhabdoviruses. Numerous myristoylation sites and an amidation site were also predicted.

G_{NS} protein

The NGAV genome contains a second glycoprotein gene immediately downstream from the G gene, separated by a single intergenic "C" nucleotide. Prior to NGAV, the only rhabdoviruses known to contain a second glycoprotein were the ephemeroviruses BEFV and ARV (Walker et al., 1991; Walker et al., 1992; Wang et al., 1994; Wang and Walker, 1993), making this an uncommon feature within the *Rhabdoviridae*. In concurrence with BEFV and ARV, this gene/protein has been designated as G_{NS} (G nonstructural). The 568-aa NGAV G_{NS} product is predicted to contain 10 *N*-glycosylation sites (unlike G which contains only four), all of which are located in the ectodomain. Like the G_{NS} proteins of BEFV and ARV, the NGAV G_{NS} protein is predicted to contain the typical characteristics of a rhabdovirus G protein, a signal peptide (residues 1 to 17), an ectodomain (residues18 to 517), an 18-aa transmembrane domain (residues 518 to 535), and a cytoplasmic tail (residues 536 to 568).

Transcription of G and G_{NS} genes varies between BEFV and ARV. In BEFV, both are transcribed as monocistronic mRNAs, whereas in ARV a disruption in the G and G_{NS} polyadenylation signals results in polycistronic mRNA transcription of G, G_{NS} and the subsequent unknown α1 and α2 ORFs (Wang and Walker, 1993). Both of the NGAV G and G_{NS} genes contain intact transcription control signals and therefore are predicted to be transcribed as monocistronic mRNAs similarly to BEFV.

There is significant similarity between the BEFV/ARV G_{NS} proteins and other rhabdovirus G proteins, suggesting that the G_{NS} proteins

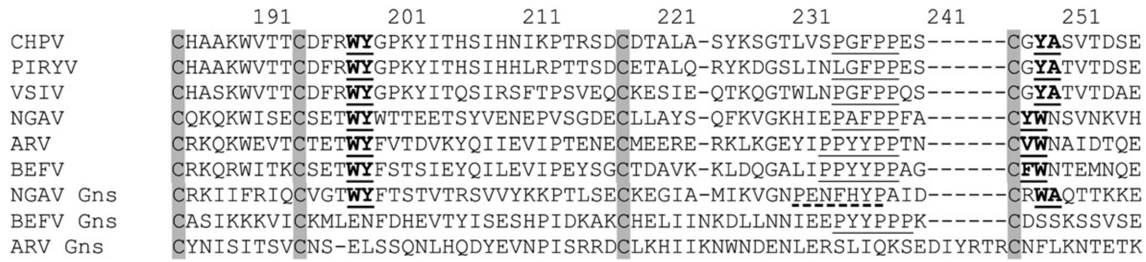


Fig. 2. ClustalW comparison of glycoprotein fusion loop regions of NGAV and other rhabdoviruses. Hydrophobic residues present at the tip of each fusion loop are highlighted in bold and underlined. The polyproline helix domain is underlined, and a potential NGAV G_{NS} polyproline helix is underlined with a dashed line. Conserved cysteine residues (L-R: C_{II}, C_{III}, C_{IV} and C_{V}) are shaded. Alignment is adapted from Johal et al. (2008).

could have resulted from a duplication of the G protein (Walker et al., 1992; Wang and Walker, 1993). Pairwise alignment of the NGAV G_{NS} protein with G proteins of NGAV, WONV, FLAV, BEFV and ARV identified high overall similarity (~40% similarity and 17–20% identity). Dotplot analyses comparing the NGAV G_{NS} and G protein sequences showed that G_{NS} lacked a ~120-aa region at the G amino-terminus (data not shown). Amino acid similarity was distributed relatively evenly across the two proteins. NGAV G_{NS} was assessed for the presence of hydrophobic residues at the tips of the fusion loops that have been shown to play a critical role in cell fusion of the VSIV G protein (Roche et al., 2006; Roche et al., 2007; Zhang and Ghosh, 1994). Equivalent residues have been identified in the BEFV and ARV G proteins, but not the G_{NS} proteins. Unlike BEFV and ARV, the NGAV G_{NS} protein contains hydrophobic residues at the tips of both loops (Fig. 2).

Cysteine residue conservation and glycoprotein structure

Rhabdovirus G proteins typically contain 10 to 18 conserved cysteine residues which form disulphide cross-links critical to their folding structure and function (Thornton, 1981; Walker and Kongsuwan, 1999). The cysteine bonding pattern appears to differ sufficiently between different genera and could be used for the taxonomic assignment of new viruses. Analysis of cysteines in the NGAV G protein showed that along with WONV and FLAV these proteins share a unique conservation pattern that is different to that of any other rhabdovirus (Table 2).

Cysteine conservation in the NGAV G_{NS} protein mirrored that of the G protein except that it contains three additional residues, one (C^{376}) seemingly unpaired within the globular domain and two (C^{421} and C^{424}) within the protein stalk. An orphaned cysteine is not usual in virus glycoproteins and such an anomaly might allow for some folding plasticity due to alternative binding. No nucleotide sequence conflicts were evident in any of the cysteine codons in the G_{NS} coding sequence, for which there was 10- to 20-fold sequencing coverage, indicating that all are authentic. It is possible that C^{376} is involved in an inter- rather than intramolecular bond and that C^{421} and C^{424} bond with each other within the stalk. A second possibility is that C^{421} and C^{424} could alternatively form a bond with C^{376}, resulting in two isomers of G_{NS}. We have tentatively designated this additional bond within NGAV G_{NS} as bond "L" in Table 2. Both NGAV G and G_{NS} contain additional cysteine residues within their predicted transmembrane regions and G also contains a cysteine within its predicted signal peptide sequence, but these are assumed to play no role in folding of the ectodomain. NGAV G and G_{NS} also share numerous conserved proline and glycine residues which, similar to previous observations of BEFV, are likely involved in forming angular turns (Chou and Fasman, 1978; Walker et al., 1992).

L protein

Common to the L polymerases of (−) ssRNA viruses are six domains (designated block I to VI) which are highly conserved (Poch et al., 1990). The domains contain motifs essential to the core functions of this enzyme. The 2105-aa NGAV L protein was compared using ClustalW to the L proteins of the selected animal rhabdoviruses BEFV, ARV, ABLV, RV, VSIV, ISFV, FLAV and WONV. This analysis suggested highest overall similarity of the NGAV L protein with the two Hart Park group viruses FLAV and WONV. The six domains that Poch et al. (1990) identified as highly conserved amongst (−) ssRNA viruses, were as expected highly conserved by all the viruses within this group, including NGAV.

The seven novel NGAV genes

Additional genes occur in some rhabdoviruses between the P-M, M-G, and G-L genes, but none have been observed between the N-P genes. Whilst in the majority of cases, additional genes are typically located at a single location (Fig. 1B), they occur at all three locations in NGAV (Fig. 1A). The seven novel NGAV genes (U1 to U7) in addition to

Table 2
Comparison of cysteine bonding patterns of NGAV G and G_{NS} proteins.

Genus/virus	Cysteine bonds											
	A $C_{I} + C_{XII}$	B $C_{II} + C_{IV}$	C $C_{III} + C_{V}$	D $C_{VI} + C_{VII}$	E $C_{VIII} + C_{X}$	F $C_{IX} + C_{XI}$	G $C_{0} + C_{XIIa}$	H $C_{Ia} + C_{VIIIb}$	I $C_{Ib} + C_{VIIIa}$	J $C_{XIIb} + C_{XIIe}$	K $C_{XIIc} + C_{XIId}$	L[a]
Hart Park group	●	●	●	●	●	●	●					
NGAV G_{NS}	●	●	●	●	●	●	●					●
Ephemerovirus	●	●	●	●	●	●	●		●	●		
BEFV G_{NS}	●	●	●		●	●	●		●			
ARV G_{NS}	●	●	●		●	●	●		●		●	
Vesiculovirus	●	●	●	●	●	●						
Lyssavirus	●	●		●	●	●		●			●	
Novirhabdovirus	●		●	●	●	●	●					
U/A Sigma virus	●	●	●		●	●						

Table adapted from previously published work with updated bonding of pairs E and F based on recently published VSV crystallisation data, as described in text. The Hart Park group includes analysis of the completely sequenced WONV, FLAV and NGAV. U/A, unassigned.

[a] L designates a new putative cysteine bonding pair found only in the NGAV G_{NS} protein as described in text.

Table 3
NGAV transcription and replication control sequences.

Junction	Gene end	IGR	Gene start
Leader–N	N.A.	N.A.	AGCAGGGATTATCCCTT*AACAG*CAAAATG
N–P	TAATTCACATGAAAAAAA	TC	AGCAGGAATCATG
P–U1	TAGTTGGACAAAACGTGAAAAAAA	TC	AGCAGTCCAATCATG
U1–U2	N.A.	N.A.	N.A.
U2–U3	CTTGAAAAAAA	TC	AGCAGGGCTCAATG
U3–M	TAACATGAAAAAAA	T	AACAGGAGTCAAGATG
M–U4	TAACATGAAAAAAA	C	AACAGGTATCAATG
U4–G	CTAGAAAAAAA	C	AACAGTTGTCAATATG
G–G_{NS}	CTTGAAAAAAA	C	AACAGTAATCAACATG
G_{NS}–U5	CATGAAAAAAAA	C	AACAGATATCATG
U5–U6	N.A.	N.A.	N.A.
U6–U7	TGACTTGAAAAAAA	C	AACAGAAATCAATATG
U7–L	TGACATGAAAAAAA	C	AACAGTTATCCATG
L–Trailer	TGA…34 nt…CTTGAAAAAAA	N.A.	N.A.

Sequences are provided in antigenome sense, as they appear on GenBank. Transcription start signals are bolded and transcription stop signals are underlined. Translation stop codons (gene end column) and start codons (gene start column) bolded and underlined. A possible alternate transcription start signal for the N protein is bolded and italicized. The G_{NS} polyadenylation signal contains an additional A nucleotide. Excessive internal untranslated sequence for the L gene end has been removed for clarity of presentation, as indicated. Stars indicate conserved nucleotides; IGR, intergenic region.

G_{NS} all possess transcription start and stop signals. The only exceptions are U1/U2 and U5/U6, which appear to be bicistronic. This indicates that mRNA transcription is likely for each of the novel ORFs (Table 3).

The predicted 139-aa U1 protein contains an *N*-glycosylation site, several phosphorylation sites and an *N*-myristoylation site. PredictProtein analysis predicted an amino-terminal ectodomain proceeded by an 18 aa transmembrane region (residues 36 to 53) and a carboxy-terminal cytoplasmic tail, but a low probability score of the transmembrane domain by the Phobius analysis indicates that its assignment is tentative.

The 133-aa U2 protein is predicted to contain numerous phosphorylation sites, a single *N*-myristoylation site and two disulfide bonds between residues C^{6}–C^{22} and C^{27}–C^{64}.

The 146-aa U3 protein is predicted to contain a single *N*-glycosylation site, several phosphorylation sites, an *N*-myristoylation site and a disulfide bond between residues C^{49}–C^{96}.

The 81-aa U4 protein (9.7 kDa) is the smallest in NGAV, contains two cysteine residues (C^{20}, C^{44}) and is predicted to contain two phosphorylation sites, an amino terminal cytoplasmic region (residues 1 to 32) and a centrally located 18-aa transmembrane domain (residues 33 to 50) proceeded by an ectodomain (residues 51 to 81). Its location between the M and G genes is synonymous to the novel small (93-aa) hydrophobic protein of the unclassified rhabdovirus TUPV which, likewise, contains a transmembrane domain (Springfeld et al., 2005).

The 114-aa U5 protein is predicted to contain several phosphorylation sites, an *N*-myristoylation site and a centrally located helix-loop-helix structure from residues 46 to 67.

The 112-aa U6 coding sequence is predicted to contain a single *N*-glycosylation site, several phosphorylation sites, a signal peptide (residues 1 to 19), an ectodomain (residue 20 to 76), a 20-aa transmembrane region (residues 77 to 96), and a 16-aa cytoplasmic tail (residue 97 to 112). A disulfide bond is predicted to form between residues C^{99} and C^{106} within the carboxy-terminal cytoplasmic region. Three regions within the signal peptide/ectodomain region of the U6 protein possess high amino acid sequence homology to the cognate region of the G_{NS} proteins (30/79 amino-terminal residues; 39%) (Fig. 3). This level of homology was not observed in any other genome regions and could be indicative of common ancestry of G_{NS} and U6.

The 153-aa U7 protein is predicted to contain two *N*-glycosylation sites, two phosphorylation sites and the majority of this protein appears to comprise a helix-sheet-helix structure.

NGAV transcription/replication control sequences

The putative transcription and replication control sequences of NGAV (Table 3) vary substantially from those of other rhabdoviruses. For the 11 mRNA species predicted to be transcribed from the NGAV genome, there are two transcription start sequence variants (AACAG or AGCAG) and four transcription stop/polyadenylation sequence variants (CATG[A_7] for N, U3, M and U7; CTTG[A_7] for U1,U2, G, U5, U6 and L; CGTG[A_7] for P; CTAG[A_7] for U4). Additionally, the G_{NS} polyadenylation signal contains eight A residues (CATG[A_8]) instead of the typical seven. In VSV, alteration of any of the first three nucleotides of the stop signal reduces transcript termination efficiency (known as "leaky" stop signals) resulting in polycistronic mRNAs, whilst altering the G immediately preceding the A_7 tract completely abolishes termination (Barr et al., 1997). Although the NGAV transcription stop signals vary considerably, the critical G

```
NGAV U6    1 MVAWWTIIILLSFKIMDFPGVIATPTLFRSMINETDLDITSRIMNYNIGP   50
              .::.|:||||......|..........:|:..|||.....:..
NGAV Gns   1 M--FYLIVILLSLVPNSKPEDHQDAPECTHVIHPYDLDDEWTEITQ----   44

NGAV U6   51 FLEKFKKFQEDLKSIWSKIKDKIEII----------KSYIIIILIIIAVIV   90
               |:....||..|||:|...:.:.::              ..|....||..:..
NGAV Gns  45 --EELNCTQEGFKSIYSGPSETVRLLYPPSSLPIRKHGYYCRKIIFRIQC   92

NGAV U6   91 VITLVTLKCFSRMITCYKSLTS*---------------------------  113
             |.|.......:|.:...|...|.
NGAV Gns  93 VGTWYFTSTVTRSVVYKKPTLSECKEGIAMIKVGNPENFHYPAIDCRWAQ  142
```

Fig. 3. Pairwise comparison of the NGAV U6 and G^{NS} proteins. Regions of high amino acid similarity are highlighted by shaded boxes. The predicted signal peptides (bold), transmembrane region (bold, underlined), *N*-glycosylation site (bold, italicized) and cysteine bond (bracketed) are highlighted. Alignment was generated using the Needleman-Wunsch algorithm.

residue is conserved in all cases, as is the first C residue. Similar variability in transcription start/stop signals occurs in the genome of WONV (Gubala et al., 2008) and variability also occurs in BEFV and ARV, but to a lesser degree (McWilliam et al., 1997; Wang and Walker, 1993). Sequence variation could be used as a regulatory mechanism for controlling relative transcript abundance (Wang and Walker, 1993). It has been shown for VSV that gene transcription is reduced by up to 30% for each transcript as the polymerase moves downstream of the genome (Iverson and Rose, 1981). Therefore "leaky" stop signals, which result in polycistronic mRNAs, could provide a means of increasing transcriptional efficiency of downstream genes. Another possibility is that the L polymerase of NGAV is less discriminatory of its transcription control sequences compared to other rhabdovirus L proteins.

As in BEFV, ARV and WONV, it appears likely that four of the novel NGAV genes (U1 and U2; U5 and U6) are almost certainly transcribed as U1/U2 and U5/U6 bicistronic mRNAs due to a lack of recognisable termination signals proceeding U1 and U5, and a lack of initiation signals preceding U2 and U6. Whilst the U2 ORF overlaps the U1 ORF by 58 nt, the U5 and U6 ORFs are separated by an 11-nt UTR (ACAAAGAAGAA).

Leader and trailer sequences

Similar to other rhabdoviruses, the 44-nt leader sequence and the 38-nt trailer sequence of NGAV display high complementarity of terminal nucleotides (14/16 nucleotides as underlined: 5′-*ACGAA*-*GAAAGAAAAGA*–) (Whelan et al., 2004). Typically animal rhabdoviruses share conservation of the three genome terminal nucleotides (5′-ACG-,–CGT-3′) which are likewise conserved in NGAV.

Protein sequence comparisons and phylogenetic analyses

From alignments of the NGAV structural protein sequences with equivalent proteins of other rhabdoviruses, it is evident that it is most similar to WONV and FLAV, followed by the ephemeroviruses. The NGAV L (60–68%) and N proteins (56–65%) displayed highest amino acid sequence similarities, followed by G (40–50%) and M proteins (42–45%). Low similarity (<20%) was evident between the P proteins, and no hits to rhabdovirus P proteins were generated by BLAST searches, highlighting their high divergence at the aa sequence level.

Phylogenetic analyses of N and G protein sequences clustered NGAV within the "Hart Park group" that contains WONV and FLAV (Figs. 4A and B). Analyses of G and G_{NS} proteins grouped the NGAV G_{NS} protein loosely with those of BEFV and ARV (Fig. 4B), which was supported by a weak bootstrap value (47/100), indicating low confidence of this branching position. As analysis of such a large set of divergent glycoproteins might distort bootstrapping of the NGAV G_{NS} protein, sequences of a restricted set of more closely related viruses (NGAV, BEFV, ARV, WONV and FLAV) were analysed. NGAV G_{NS} branching occurred at the same position with a slightly increased but still low bootstrap value (58/100) (data not shown). Additional analyses performed using different algorithms gave varying values (Fitch, 51; Kitch, 97; Protpars, 100). Kitch generated the most consistent result and a high supporting bootstrap value for the branching of NGAV G_{NS} with the BEFV and ARV G_{NS} proteins. Analyses performed using the maximum likelihood method similarly indicated that NGAV G_{NS} branched with the BEFV/ARV G_{NS} proteins. However, this association will remain tentative until more rhabdovirus G_{NS}-like proteins are discovered and compared.

Protein synthesis in infected cells

A study to detect NGAV proteins synthesised over time in infected cells was performed by analysing proteins on SDS-PAGE and Western blotting using an anti-NGAV MAF (results not shown). The analyses suggested the presence of eight virus-specific protein bands. Four of the most prominent bands were consistent in size with the structural proteins G (81 kDa), N (48 kDa), P (26 kDa) and M (23 kDa). Although obscured by the 81-kDa G protein, a faint band corresponding approximately in size to the G_{NS} protein (65 kDa) also appeared from 16 h onwards. The three remaining bands were larger than the putative G band, and could correspond to different glycosylation states of G, which has previously been observed for BEFV, ARV, FLAV and WONV (Boyd and Whitaker-Dowling, 1988; Gubala et al., 2008; Walker et al., 1991). No bands corresponding to the seven small novel proteins (U1 to U7) were detected, nor was the 242-kDa L protein.

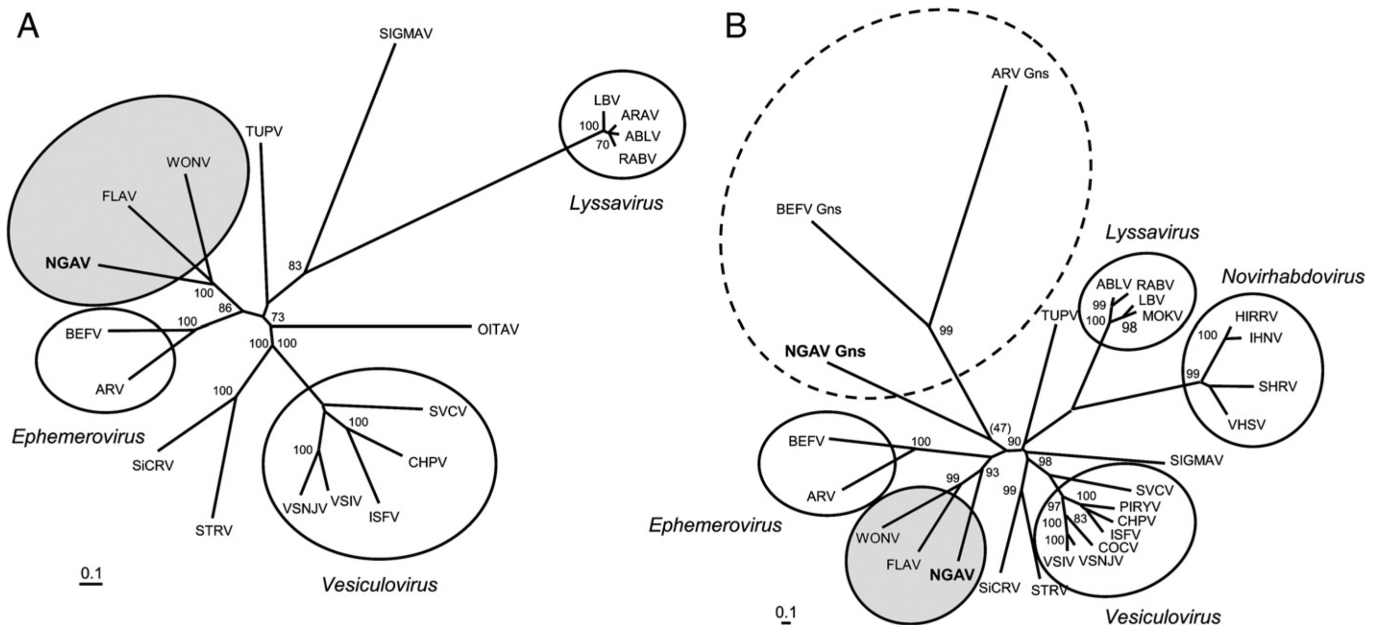


Fig. 4. Phylogenetic analysis of selected rhabdovirus full-length N (A) and G (B) proteins. Viruses from the genera *Cytorhabdovirus* and *Nucleorhabdovirus* (N and G) and *Novirhabdovirus* (N), which infect plants and fish, have not been included for clarity. The shaded areas on both trees highlight NGAV and the "Hart Park" cluster. The G_{NS} proteins of NGAV, BEFV and ARV are highlighted by a dashed circle (B). Bootstrap values greater than 70 (from 100) have been indicated. Bar represents expected substitutions per site.

Prevalence of NGAV in wildlife and livestock

A set of 1657 sera collected from various animals in the Northern Territory, Australia, between 1993 and 2007 was tested for exposure to NGAV using a virus neutralization test (VNT), as summarised in Table 4. Following the initial screening of all sera diluted 1:5, 267 potential positives were subsequently titrated at dilutions of 1:10, 1:20, 1:40 and 1:80. Following titration, 106/267 sera were scored as positive at a titre ≥10. The majority of these sera were from macropods: 88 agile wallaby (*Macropus agilis*), 16 black wallaroo (*Macropus robustus woodwardii*) and 1 kangaroo (*Macropus rufus*). A significant proportion possessed titres≥64 (46 wallabies and 2 wallaroo), and of these, 18 had titres between 160 and 640. Neutralization at titre of 10 also occurred with serum from a buffalo, and low-level neutralization occurred with sera from three other buffalo and two sentinel cows. In summary, 20% of the wallaby, 34% of the wallaroo and 5% of the kangaroo sera tested were positive for NGAV neutralizing antibodies.

Discussion

The conventional molecular evolution theory of purifying selection suggests there should be preference for smaller and less complex genomes. However, it is evident from the growing number of complete genome sequences of rhabdoviruses (Basurco and Benmansour, 1995; Dietzgen et al., 2006; Gubala et al., 2008; Huang et al., 2003; Reed et al., 2005; Revill et al., 2005; Scholthof et al., 1994; Springfeld et al., 2005; Tanno et al., 2000; Walker et al., 1992; Wang et al., 1994; Wang and Walker, 1993) that they are preserving numerous additional genes, indicating that rather than hindrance, these genes likely confer unique evolutionary advantages. Conversely, they could also be kept or acquired by the virus out of strict necessity. Until recently, the observed location of these additional genes was only between the G and L or the P and M genes. Recently, the rhabdovirus TUPV was found to contain a novel gene between M and G, and WONV was found to contain novel genes in three different genome locations (Gubala et al., 2008; Springfeld et al., 2005). Here we report the complete genome sequence of NGAV, which shows it to contain eight putative additional genes, three between P and M, one between M and G, and four between G and L, resulting in the most complex genome organisation of any rhabdovirus studied to date.

In VSV, gene transcript levels progressively decrease by approximately 25% to 30% as the polymerase transcribes each downstream gene and it has been found that the presence of additional genes between N and P disrupts a critical N:P protein ratio which impairs viral replication (Iverson and Rose, 1981; Wertz et al., 2002). This finding likely explains why there are no rhabdoviruses known to contain a functional gene between N and P and why locating additional genes between P and M provides the first opportunity for high-level transcription without dire consequences for virus replication. NGAV along with WONV and FLAV in the "Hart Park" phylogenetic cluster are the only animal rhabdoviruses yet to be found with additional genes between P and M. It is interesting that the plant-infecting nucleorhabdoviruses and cytorhabdoviruses also contain genes at this location (Dietzgen et al., 2006; Heaton et al., 1989; Reed et al., 2005; Revill et al., 2005; Scholthof et al., 1994; Tanno et al., 2000). Although there is no phylogenetic data cementing a clear link between the very distantly related Hart Park viruses and the plant rhabdoviruses, the presence of vastly diverse genes between P and M are a prominent commonality that perhaps represent evolutionary relics that link the animal and plant viruses. It is also possible that these genes arose opportunistically through a common process, such as gene duplication.

Of all the additional genes present in the NGAV genome, perhaps the most intriguing is the second glycoprotein gene (G_{NS}) located immediately downstream of the G protein gene. The only other rhabdoviruses known to contain a G_{NS} gene are the two closely related ephemeroviruses BEFV and ARV. Comparisons of BEFV and ARV G and G_{NS} sequences indicate that it is likely G_{NS} arose due to a gene duplication event as their G_{NS} genes appear to have evolved separately from the G genes, being more closely related to each other than to other rhabdovirus G proteins. It was hypothesised, therefore, that these two viruses arose from a common ancestor which contained two related glycoprotein genes (Wang and Walker, 1993). The two G_{NS} proteins, although clearly related, share less sequence homology than the two G proteins, suggesting that the G_{NS} gene has diverged more rapidly than the G gene (Wang and Walker, 1993). Although phylogenetically the NGAV G_{NS} protein sequence is more closely related to the BEFV/ARV G_{NS} proteins than to G proteins of other rhabdoviruses, bootstrap values are low and more G_{NS}-related proteins will be needed to confirm this closer apparent relationship. However, low bootstrap values are not unusual for rhabdovirus proteins due to their wide evolutionary divergence and values of less than 50 for the L proteins of the unassigned rhabdoviruses TUPV, Oita virus, and trout rhabdovirus are a good example of this diversity (Springfeld et al., 2005). The analyses here show the NGAV G_{NS} protein to be more distantly related than are the BEFV and ARV G_{NS} proteins but less divergent from rhabdovirus G proteins. However, the G_{NS} protein sequence relationships are so tenuous that it is currently impossible to hypothesise whether they originated from a common ancestor of NGAV and the ephemeroviruses.

The function and significance of the second glycoprotein in rhabdoviruses remains unknown and requires further analysis. Despite their numerous similarities, the BEFV G and G_{NS} proteins are antigenically distinct (Hertig et al., 1995; Johal et al., 2008). They share significant amino acid similarity (36%) and several structural characteristics dictated by conserved cysteine, proline and glycine residues (Walker et al., 1992). Despite these similarities, the BEFV G_{NS} protein fails to react with monoclonal antibodies or polyclonal antisera raised to the G protein (Hertig et al., 1995; Johal et al., 2008). Furthermore, unlike G protein antisera, polyclonal antisera raised to G_{NS} do not neutralize the virus and do not protect cattle against BEFV challenge (Hertig et al., 1995). These data indicate a considerable difference in antigenic determinants of the two proteins.

Like most viral proteins, it is likely G_{NS} has several roles. In BEFV, G_{NS} is expressed constitutively with G in both mammalian and insect cell culture (Johal et al., 2008; Walker et al., 1992). Although it localises to the cell surface in both mammalian and insect cells, data indicate that it is not incorporated into virions, at least in the former. Whether the G_{NS} protein of BEFV, ARV and NGAV mediates some viral replication function or directs cellular or immune responses in infected tissues in the mammalian host or insect vector remain to

Table 4
Serum samples analysed by virus neutralization test for exposure to NGAV.

Animal species	No. of samples	No. of positive titre ≥10	No. of positive titre ≥64
Cattle	275	0 (30)	–
Buffalo	296	1 (39)	–
Pig	283	0 (24)	–
Bird (wild)	211	0 (10)	–
Bird (farmed)	28	0 (1)	–
Wallaby	437	88 (132)	46
Wallaroo	47	16 (21)	2
Kangaroo	19	1 (2)	0
Horse	23	0 (6)	–
Rat	17	0 (1)	–
Python	21	0 (1)	–
Total	1657	106 (267)	48

Numbers in brackets represent the number of positive samples that were titrated following the initial screening (1:5 dilution).

be determined. A notable difference between the NGAV G_{NS} protein and the ephemerovirus G_{NS} proteins is that it contains the appropriate hydrophobic residues at the tips of the fusion loops that are likely involved in cell fusion. Cell fusion by rhabdovirus glycoproteins is caused by pH-induced conformational changes of fusion peptides which consequently become exposed to interact with lipid membranes (Fredericksen and Whitt, 1995; Roche et al., 2006; Roche et al., 2007; Whitt et al., 1990; Zhang and Ghosh, 1994). VSIV G protein contains two putative fusion loops comprising hydrophobic residues (Roche et al., 2006; Roche et al., 2007; Zhang and Ghosh, 1994). Whilst BEFV G protein likewise can cause cell fusion that is pH-induced, its G_{NS} protein lacks this function, which is proposed to be due to the absence of these critical residues (Johal et al., 2008). The presence of these residues in NGAV G_{NS} suggests that it could have the potential to induce cell fusion, assuming that the loop composition is a reliable marker of this activity.

Whilst it has been proposed that the small novel proteins of BEFV and ARV might play some role in regulating transcription and replication (Dhillon et al., 2000), it is possible they perform several functions in the vector/host. Small proteins characteristic of viroporins occur in BEFV, ARV, WONV and FLAV (Gubala et al., 2008; McWilliam et al., 1997; Wang et al., 1994), and TUPV also contains a small hydrophobic (SH) protein with characteristics similar to viroporins (Springfeld et al., 2005). Viroporins are small proteins that occur in diverse viruses and form pores in host cell membranes enabling the passage of ions and small molecules (Gonzalez and Carrasco, 2003). They are believed to be nonessential but to significantly enhance virus replication. Although they possess little amino acid similarity to each other, the putative viroporins of BEFV, ARV, WONV and FLAV all contain a central transmembrane domain surrounded by charged residues as well as a highly basic domain at the carboxy terminus (Gubala et al., 2008; McWilliam et al., 1997; Wang et al., 1994). Compared with these putative rhabdovirus viroporins, the three novel NGAV proteins with transmembrane domains (U1, U4 and U6) are similar in size but none contain a highly basic carboxy terminus. Moreover, unlike viroporins, which typically contain long cytoplasmic tails (Gonzalez and Carrasco, 2003), the NGAV U6 protein transmembrane domain is located just 16 aa upstream of the carboxy terminus. However, the characteristics of viroporins remain speculative and thus it cannot be precluded that these NGAV proteins possess similar functions.

There are several mechanisms by which new genes might arise in rhabdoviruses. Since G_{NS} proteins are clearly related to the G proteins, it has been suggested that the G_{NS} gene could have arisen through duplication of the upstream gene due to the loose association of RNA polymerase with the template RNA and similar in mechanism to how defective interfering genomes are formed (Lazzarini et al., 1981; Perrault, 1981; Walker et al., 1992). An alternate theory is that it might have been acquired during a mixed infection with two rhabdoviruses (Walker et al., 1992). However, this seems less likely as recombination between different rhabdoviruses is believed to be a low probability event (Domingo and Holland, 1997). The mechanism by which several rhabdoviruses have acquired multiple small novel genes that are not clearly related is not obvious. It is possible that they arose through polymerase-mediated duplication events and have subsequently diverged beyond recognition. Why they have occurred in some rhabdoviruses like NGAV to generate larger and more complex genome structures, whilst other viruses have maintained a minimal genome comprising only five essential genes, remains to be explored. Currently there is insufficient data to know whether genome complexity of some rhabdoviruses existing today has increased from ancestral viruses but it seems likely that they have evolved from dynamic populations with a capacity to either gain or lose genes in response to selective pressures and environmental adaptation.

Since functions of the small novel proteins of rhabdoviruses are not clear, it is difficult to predict whether they are expressed in high or low amounts, or whether their expression is species and cell type specific. Without the availability of specific antisera, their detection in cell culture is difficult. For example, whilst there is some data suggesting the low-level expression of the novel β protein of ARV and several of the novel proteins of WONV, the three novel proteins of BEFV have never been detected despite extensive efforts (Gubala et al., 2008; Walker et al., 1991; Wang and Walker, 1993). This study has suggested that none of the novel NGAV proteins were detected in western blotting experiments using polyclonal MAF. This is not surprising because the detection of the novel NGAV proteins may be facing the same types of challenges as BEFV and ARV, which are as yet undefined. In addition, there are limitations to using a nonspecific, uncharacterised antiserum (the MAF) to detect viral proteins. It is not possible to know whether the novel proteins were present during the generation of the MAF (the MAF was generated to a crude preparation of infected suckling mouse brain). It is also possible that the novel proteins have low immunogenicity, or they may not be expressed at all or may be expressed at very low levels in mammalian cell culture. As such, these results are not conclusive but they should be taken into consideration in subsequent studies. It would be most beneficial to produce antisera specific to each of the novel proteins in the future.

When first isolated in 1970, a limited serologic survey suggested that macropods and cattle were potential hosts of NGAV (Doherty et al., 1973). Examination of sera collected opportunistically at wildlife parks and care centres in northern Australia showed that NGAV is still circulating in wallabies, wallaroos and kangaroos, which might have implications for the health of such macropods. A means of regular monitoring of these macropods is needed to establish whether NGAV is associated with disease, and molecular detection tests devised from genome sequence information presented here should prove useful in such monitoring.

Phylogenetic analyses of rhabdovirus N and G proteins and comparisons of the other structural proteins show that a distinct cluster is formed by NGAV, WONV and FLAV. Predictive modelling for various rhabdovirus G proteins (Walker, 2008; Walker and Kongsuwan, 1999), and more recently crystallization data for the VSV G protein (Roche et al., 2006; Roche et al., 2007) have afforded insights into cysteine linkage patterns as potentially useful markers in the taxonomic assignment of new viruses. The shared cysteine bonding pattern of the NGAV, WONV and FLAV G proteins is clearly different to other rhabdoviruses (Walker and Kongsuwan, 1999). Taken together with the phylogenetic data, this supports the formation of a new genus comprising NGAV, WONV and FLAV for which complete genomic sequences are now available. However, placing NGAV with WONV and FLAV is complicated by the presence of the second glycoprotein gene (G_{NS}), which to date has only been detected in viruses in the genus *Ephemerovirus*. It is apparent that as the genomes of rhabdoviruses such as NGAV are sequenced, more complexity will be encountered which taxonomy will need to accommodate. For this reason and to improve virus detection methods to better assess potential disease threats as well as develop recombinant vaccines, there is a clear need to determine the genome sequences of the many yet uncharacterised rhabdoviruses and determine the functional role of the novel accessory genes present in viruses such as NGAV.

Materials and methods

Virus propagation and RNA extraction

NGAV (strain MRM14556) and BEFV (strain CS1927) were propagated in BHK-BSR cells grown in basal medium Eagle supplemented with 10 mM HEPES, 6.7 mM $NaHCO_3$, 2 mM L-glutamine, 137 μM streptomycin, 80 U/ml penicillin and 5% (growth media, GM) or 2.5% (maintenance media, MM) fetal calf serum, at 37 °C. At 4 days

postinfection (dpi), the infected cell culture supernatant was centrifuged at 2000 × g for 10 min to remove cellular debris and a virus pellet was obtained by ultracentrifugation at 206,000 × g for 40 min in a Beckman type 55.2Ti rotor. Total RNA was extracted from the crude virus pellet using the RNeasy Mini Kit (QIAGEN, Germany) and quantified using the GeneQuant II RNA/DNA calculator.

PCR-select cDNA subtraction to obtain NGAV genomic sequence data

NGAV and BEFV were used as the tester and driver respectively in PCR-select cDNA subtraction reactions. We have previously modified this method to obtain overlapping genome fragments for increased efficiency by using three different restriction enzymes (Gubala et al., 2008). Here, the method was further modified by pooling RNA from three different rhabdoviruses (genome sequence analysis of the other two viruses is not described here). The genome fragments generated were cloned into pCR-Blunt II-TOPO and sequenced as previously described (Gubala et al., 2008). Additional sequences were also obtained by 454 Life Sciences sequencing (Roche) analyses of uncloned DNA obtained from secondary PCR amplification of the PCR-select cDNA subtraction reactions. Sequences generated from these two approaches were used to generate a single contiguous sequence of the NGAV genome. For confirmation, the contig was used to design primer pairs to amplify sequences (except for the genome termini) spanning the entire NGAV genome as described previously (Gubala et al., 2008).

Characterisation of genome extremities by RACE

A rapid amplification of cDNA ends (RACE) method was used to obtain the terminal genome sequences as previously described (Gubala et al., 2008) except that RNA was extracted directly from 100 µl infected cell culture supernatant rather than from virus concentrated by ultracentrifugation. The 3′ RACE method used the primers 5′-GAGGTGCATCCAGCTATTTG-3′ (located 574 nt from 3′ terminus) and 5′-CATAGCGGATAGCGTGTGCTAC-3′ (located 284 nt from the 3′ terminus) in the primary and secondary PCR, respectively. The 5′ RACE method used the primers 5′-CATTGGTATAATCAGCTG-GATTGC-3′ (located 441 n from the 5′ terminus) for cDNA synthesis and 5′-TTGCCACTGGAAACACTC-3′ (located 415 nt from the 5′ terminus) and 5′-CACACTTATGAGTCGATAG-3′ (located 386 nt from 5′ terminus) in the primary and secondary PCR, respectively. To increase specificity, the RACE adaptor-specific primer used previously (19) was modified to contain three additional nucleotides (5′-AACGCCATTTCCACCTTCTCTTC*ACG*-3′) homologous to the conserved genome termini of animal rhabdoviruses. The PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN) and sequenced directly using the virus-specific primers.

DNA sequencing and sequence analyses

Sequencing was performed using the BigDye Terminator v. 1.1 kit and the ABI PRISM 3130*xl* Genetic Analyzer (both from Applied Biosystems, USA) and the 454 Life Sciences Sequencer (Roche) according to the manufacturer's protocols.

Sequence assembly and routine sequence management were performed using the programs SeqMan Pro v. 8.0.2 (Lasergene v. 8, DNASTAR), Clone Manager v. 9 (Sci Ed Central) and Artemis (Altschul et al., 1997). Sequence similarity searches of GenBank/EMBL and SWISSPROT databases were performed using BLAST (http://www.ncbi.nlm.nih.gov) (Altschul et al., 1997). Amino acid sequence identity levels of deduced NGAV proteins to equivalent proteins of other members of the *Rhabdoviridae* were determined from ClustalW [Accurate] pairwise alignments using MegAlign v. 7.2.1 (Lasergene). Protein sequences were analysed using PredictProtein (http://www.predictprotein.org) (Rost et al., 2004), ProtScale and SignalP programs available at the ExPASy Proteomics Server (http://au.expasy.org) and Phobius (http://phobius.cbr.su.se/).

Phylogenetic analyses were performed using programs available through BioManager at the ANGIS web interface (http://www.angis.org.au). ClustalW alignments of selected proteins were edited manually at both ends to avoid gaps and ensure maximum compatibility of the sequences compared. Evolutionary relationships were estimated using the distance matrix method (Protdist) and neighbour-joining phylogenetic trees were drawn using TreeView v. 1.6.6 (Page, 1996). Bootstrap resampling (100 replicates) was performed using Seqboot to confirm the statistical reliability of the phylogenetic trees. Dotplot analyses were performed using the Dotmatcher program available at ANGIS and employed a window of 15 and a threshold of 30 using the EBLOSUM 62 matrix.

Immunoblotting

BHK-BSR cell monolayers were grown to 80% confluency in 10 cm^2 tissue culture plates (Nunc) and were either mock infected or infected with NGAV at a multiplicity of 10 $TCID_{50}$/cell. The inoculum was removed after 1 h and replaced with fresh MM. At 8, 16, 24 and 32 hpi the cells were harvested into phosphate-buffered saline A (PBSA) containing 1 × Complete Protease Inhibitor Cocktail (Roche), followed by a wash and resuspension in 100 µl of this solution. The cell preparations (20 µl) were then treated according to the Invitrogen instructions for NuPAGE Novex SDS-PAGE gel analysis of proteins and were analysed in NuPAGE Novex 12% Bis-Tris SDS-PAGE gels. Proteins were transferred onto Hybond-LFP PVDF membrane (Amersham) according to the Invitrogen western blotting transfer protocol and were treated with immune polyclonal mouse ascitic fluid (MAF) (1/200) raised to NGAV prepared according to the method previously described (Sartorelli et al., 1966), followed by binding to sheep antimouse HRP-conjugated Ig (1/1000) (Silenus, Australia). Membranes were immersed in Western Lightning Chemiluminescence Substrate Reagent Plus (PerkinElmer Life Sciences) for 1 min for development of chemiluminescence and exposed to Kodak BioMax Light film. Anti-NGAV MAF was produced by inoculating mice with a heat-inactivated infected suckling mouse brain at the CSIRO Long Pocket Laboratories, Brisbane. This MAF has been used previously in complement fixation, immunofluorescence and virus neutralization assays and does not cross-react with any other Australian rhabdovirus.

Virus neutralization tests

Serum samples from various animal species were screened in a virus neutralization test (VNT) to assess for infection by NGAV. Initially, serum samples were diluted 1:5 in cell culture medium and duplicate aliquots of 50 µl were placed into 96-well flat-bottomed microplates. Cell culture medium only (serum toxicity control) or medium pretitrated to contain 100 $TCID_{50}$ of virus (50 µl), were added to each well and incubated at 37 °C with 5% CO_2 for 1 h. A 100 µl cell suspension containing 2×10^5 BHK-BSR cells per ml was added to each well and the plates were incubated under the same conditions. To determine the optimal day for the reading of plates positive serum control wells containing anti-NGAV MAF were concurrently set up. Plates were scored for inhibition of CPE at 5 dpi and a two-fold dilution series (1:10, 1:20. 1:40 and 1:80) of each positive serum was subsequently performed in duplicate using the same procedure. Serum neutralization titres were calculated using the Reed and Muench 50% end point method (Reed and Muench, 1938).

Sequence database accession numbers

The NGAV genome sequence has been deposited in GenBank and has the accession number FJ715959. The accession numbers of other

viral sequences used in comparisons and phylogenetic analyses were: ABLV, *Australian bat lyssavirus* (NC_003243); ARAV, Aravan virus (EF614259); ARV, *Adelaide River virus* (U10363, L09207, U05987); BEFV, *Bovine ephemeral fever virus* (NC_002526); CHPV, *Chandipura virus* (AY614724, AY614717); COCV, *Cocal virus* (AF045556); FLAV, Flanders virus (AH012179); HIRRV, *Hirame rhabdovirus* (NC_005093); IHNV, *Infectious hematopoietic necrosis virus* (NC_001652); ISFV, *Isfahan virus* (AJ810084); LBV, *Lagos bat virus* (DQ499944, AF298148); MOKV, *Mokola virus* (NC_006429); OITAV, Oita virus (AB116386); PIRYV, *Piry virus* (D26175); RABV, *Rabies virus* (NC_001542); STRV, Sea trout rhabdovirus (AF434992); SIGMAV, (X91062); SiCRV, Siniperca chuatsi rhabdovirus (NC_008514); SHRV, *Snakehead rhabdovirus* (NC_000903); SVCV, Spring viremia of carp virus (NC_002803); TUPV, Tupaia virus (NC_007020); VSIV, *Vesicular stomatitis Indiana virus* (NC_001560); VSNJV, *Vesicular stomatitis New Jersey virus* (K02379,V01214); VHSV, *Viral hemorrhagic septicaemia virus* (NC_000855); and WONV, Wongabel virus (NC_011639).

Acknowledgments

We would like to acknowledge R. L Doherty and colleagues for collecting midges and isolating NGAV. We gratefully thank Dr Jeff Cowley for reviewing this manuscript, his expert advice and valuable help throughout this work. Sincere thanks goes to Tony Pye for providing the sequencing service and we acknowledge the trial sequencing service provided by 454 Life Sciences. We thank Dr Dieter Bulach for assembly of sequences generated from the 454 sequencing and advice with bioinformatics. We acknowledge the expert technical assistance of Christine Rioux and Rachel Amos-Ritchie. We gratefully acknowledge the funding contributed by the Australian Biosecurity CRC in support of this work.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17), 3389–3402.
Barr, J.N., Whelan, S.P., 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. J. Virol. 71 (11), 8718–8725.
Basurco, B., Benmansour, A., 1995. Distant strains of the fish rhabdovirus VHSV maintain a sixth functional cistron which codes for a nonstructural protein of unknown function. Virology 212 (2), 741–745.
Bourhy, H., Cowley, J.A., Larrous, F., Holmes, E.C., Walker, P.J., 2005. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. J. Gen. Virol. 86 (10), 2849–2858.
Boyd, K.R., Whitaker-Dowling, P., 1988. Flanders virus replication and protein synthesis. Virology 163 (2), 349–358.
Calisher, C.H., Karabatsos, N., Zeller, H., Digoutte, J.P., Tesh, R.B., Shope, R.E., Travassos da Rosa, A.P., St George, T.D., 1989. Antigenic relationships among rhabdoviruses from vertebrates and hematophagous arthropods. Intervirology 30 (5), 241–257.
Carley, J.G., Standfast, H.A., Kay, B.H., 1973. Multiplication of viruses isolated from arthopods and vertebrates in Australia in experimentally infected mosquitoes. J. Med. Entomol. 10 (3), 244–249.
Chen, B.J., Lamb, R.A., 2008. Mechanisms for enveloped virus budding: can some viruses do without an ESCRT? Virology 372 (2), 221–232.
Chou, P.Y., Fasman, G.D., 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–148.
Coll, J.M., 1995. The glycoprotein G of rhabdoviruses. Arch. Virol. 140 (5), 827–851.
Crysler, J.G., Lee, P., Reinders, M., Prevec, L., 1990. The sequence of the nucleocapsid protein (N) gene of Piry virus: possible domains in the N protein of vesiculoviruses. J. Gen. Virol. 71 (9), 2191–2194.
Das, S.C., Pattnaik, A.K., 2005. Role of the hypervariable hinge region of phosphoprotein P of vesicular stomatitis virus in viral RNA synthesis and assembly of infectious virus particles. J. Virol. 79 (13), 8101–8112.
de Hoop, M.J., Ab, G., 1992. Import of proteins into peroxisomes and other microbodies. Biochem. J. 286 (3), 657–669.
Dhillon, J., Cowley, J.A., Wang, Y., Walker, P.J., 2000. RNA polymerase (L) gene and genome terminal sequences of ephemeroviruses bovine ephemeral fever virus and Adelaide River virus indicate a close relationship to vesiculoviruses. Virus Res. 70 (1–2), 87–95.
Dietzgen, R.G., Callaghan, B., Wetzel, T., Dale, J.L., 2006. Completion of the genome sequence of Lettuce necrotic yellows virus, type species of the genus Cytorhabdovirus. Virus Res. 118 (1–2), 16–22.
Doherty, R.L., Carley, J.G., Standfast, H.A., Dyce, A.L., Kay, B.H., Snowdon, W.A., 1973. Isolation of arboviruses from mosquitoes, biting midges, sandflies and vertebrates collected in Queensland, 1969 and 1970. Trans. R. Soc. Trop. Med. Hyg. 67 (4), 536–543.
Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51, 151–178.
Fredericksen, B.L., Whitt, M.A., 1995. Vesicular stomatitis virus glycoprotein mutations that affect membrane fusion activity and abolish virus infectivity. J. Virol. 69 (3), 1435–1443.
Gonzalez, M.E., Carrasco, L., 2003. Viroporins FEBS Lett. 552 (1), 28–34.
Gubala, A.J., Proll, D.F., Barnard, R.T., Cowled, C.J., Crameri, S.G., Hyatt, A.D., Boyle, D.B., 2008. Genomic characterisation of Wongabel virus reveals novel genes within the Rhabdoviridae. Virology 376 (1), 13–23.
Harty, R.N., Paragas, J., Sudol, M., Palese, P., 1999. A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding. J. Virol. 73 (4), 2921–2929.
Heaton, L.A., Hillman, B.I., Hunter, B.G., Zuidema, D., Jackson, A.O., 1989. Physical map of the genome of sonchus yellow net virus, a plant rhabdovirus with six genes and conserved gene junction sequences. Proc. Natl. Acad. Sci. U. S. A. 86 (22), 8665–8668.
Hertig, C., Pye, A.D., Hyatt, A.D., Davis, S.S., McWilliam, S.M., Heine, H.G., Walker, P.J., Boyle, D.B., 1995. Vaccinia virus-expressed bovine ephemeral fever virus G but not G(NS) glycoprotein induces neutralizing antibodies and protects against experimental infection. J. Gen. Virol. 77 (4), 631–640.
Hooper, P.T., Lunt, R.A., Gould, A.R., Hyatt, A.D., Russell, G.M., Kattenbelt, J.A., Blacksell, S.D., Reddacliff, L.A., Kirkland, P.D., Davis, R.J., Durham, P.J., Bishop, A.L., Waddington, J., 1999. Epidemic of blindness in kangaroos—evidence of a viral aetiology. Aust. Vet. J. 77 (8), 529–536.
Huang, Y., Zhao, H., Luo, Z., Chen, X., Fang, R.X., 2003. Novel structure of the genome of Rice yellow stunt virus: identification of the gene 6-encoded virion protein. J. Gen. Virol. 84 (8), 2259–2264.
Irie, T., Nagata, N., Yoshida, T., Sakaguchi, T., 2008. Paramyxovirus Sendai virus C proteins are essential for maintenance of negative-sense RNA genome in virus particles. Virology 374 (2), 495–505.
Iverson, L.E., Rose, J.K., 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell 23 (2), 477–484.
Johal, J., Gresty, K., Kongsuwan, K., Walker, P.J., 2008. Antigenic characterization of bovine ephemeral fever rhabdovirus G and GNS glycoproteins expressed from recombinant baculoviruses. Arch. Virol. 153 (9), 1657–1665.
Karabatsos, N. (Ed.), 1985. The American Committee on Arthropod-borne Viruses: International Catalog of Arboviruses Including Certain Other Viruses of Vertebrates, 3rd ed. American Society for Tropical Medicine and Hygiene, San Antonio, Texas.
Kay, B.H., Boreham, P.F.L., Dyce, A.L., Standfast, H.A., 1978. Blood feeding of biting midges (Diptera: Ceratopogonidae) at Kowanyama, Cape York Peninsula, North Queensland. J. Aust. ent Soc. 17, 145–149.
Kouznetzoff, A., Buckle, M., Tordo, N., 1998. Identification of a region of the rabies virus N protein involved in direct binding to the viral RNA. J. Gen. Virol. 79 (5), 1005–1013.
Kretzschmar, E., Peluso, R., Schnell, M.J., Whitt, M.A., Rose, J.K., 1996. Normal replication of vesicular stomatitis virus without C proteins. Virology 216 (2), 309–316.
Lazzarini, R.A., Keene, J.D., Schubert, M., 1981. The origins of defective interfering particles of the negative-strand RNA viruses. Cell 26 (2), 145–154.
Luo, Z.L., Fang, R.X., 1998. Structure analysis of the rice yellow stunt rhabdovirus glycoprotein gene and its mRNA. Arch. Virol. 143 (12), 2453–2459.
Mavrakis, M., McCarthy, A.A., Roche, S., Blondel, D., Ruigrok, R.W., 2004. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. J. Mol. Biol. 343 (4), 819–831.
McWilliam, S.M., Kongsuwan, K., Cowley, J.A., Byrne, K.A., Walker, P.J., 1997. Genome organization and transcription strategy in the complex GNS-L intergenic region of bovine ephemeral fever rhabdovirus. J. Gen. Virol. 78 (6), 1309–1317.
Nagai, Y., Kato, A., 2004. Accessory genes of the paramyxoviridae, a large family of nonsegmented negative-strand RNA viruses, as a focus of active investigation by reverse genetics. Curr. Top Microbiol. Immunol. 283, 197–248.
Page, R.D., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12 (4), 357–358.
Peluso, R.W., Richardson, J.C., Talon, J., Lock, M., 1996. Identification of a set of proteins (C' and C) encoded by the bicistronic P gene of the Indiana serotype of vesicular stomatitis virus and analysis of their effect on transcription by the viral RNA polymerase. Virology 218 (2), 335–342.
Perrault, J., 1981. Origin and replication of defective interfering particles. Curr. Top Microbiol. Immunol. 93, 151–207.
Poch, O., Blumberg, B.M., Bougueleret, L., Tordo, N., 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. 71 (5), 1153–1162.
Reddacliff, L., Kirkland, P., Philbey, A., Davis, R., Vogelnest, L., Hulst, F., Blyde, D., Deykin, A., Smith, J., Hooper, P., Gould, A., Hyatt, A., 1999. Experimental reproduction of viral chorioretinitis in kangaroos. Aust. Vet. J. 77 (8), 522–528.
Reed, L.J., Muench, H., 1938. A simple method of estimating fifty percent endpoints. Am. J. Hyg. 27, 493–497.
Reed, S.E., Tsai, C.W., Willie, K.J., Redinbaugh, M.G., Hogenhout, S.A., 2005. Shotgun sequencing of the negative-sense RNA genome of the rhabdovirus Maize mosaic virus. J. Virol. Methods 129 (1), 91–96.
Revill, P., Trinh, X., Dale, J., Harding, R., 2005. Taro vein chlorosis virus: characterization and variability of a new nucleorhabdovirus. J. Gen. Virol. 86 (2), 491–499.
Ribeiro Jr., E.A., Favier, A., Gerard, F.C., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R.W., Blackledge, M., Jamin, M., 2008. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. J. Mol. Biol. 382 (2), 525–538.
Roche, S., Bressanelli, S., Rey, F.A., Gaudin, Y., 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science 313 (5784), 187–191.

Roche, S., Rey, F.A., Gaudin, Y., Bressanelli, S., 2007. Structure of the perfusion form of the vesicular stomatitis virus glycoprotein G. Science 315 (5813), 843–848.
Rost, B., Yachdav, G., Liu, J., 2004. The PredictProtein server. Nucleic Acids Res. 32, W321–W326 (Web Server issue).
Sartorelli, A.C., Fischer, D.S., Downs, W.G., 1966. Use of sarcoma 180/TG to prepare hyperimmune ascitic fluid in the mouse. J. Immunol. 96 (4), 676–682.
Scholthof, K.B., Hillman, B.I., Modrell, B., Heaton, L.A., Jackson, A.O., 1994. Characterization and detection of sc4: a sixth gene encoded by sonchus yellow net virus. Virology 204 (1), 279–288.
Spiropoulou, C.F., Nichol, S.T., 1993. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. J. Virol. 67 (6), 3103–3110.
Springfeld, C., Darai, G., Cattaneo, R., 2005. Characterization of the Tupaia rhabdovirus genome reveals a long open reading frame overlapping with P and a novel gene encoding a small hydrophobic protein. J. Virol. 79 (11), 6781–6790.
Tanno, F., Nakatsu, A., Toriyama, S., Kojima, M., 2000. Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. Arch. Virol. 145 (7), 1373–1384.
Thornton, J.M., 1981. Disulphide bridges in globular proteins. J. Mol. Biol. 151 (2), 261–287.
Tordo, N., Poch, O., Ermine, A., Keith, G., Rougeon, F., 1986. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? Proc. Natl. Acad. Sci. U. S. A. 83 (11), 3914–3918.
Walker, P.J., 2008. Bovine ephemeral fever virus, In: Mahy, B.W.J., Van Regenmortel, M. H.V. (Eds.), Encyclopedia of virology, 3rd ed. Elsevier Academic Press, pp. 354–362.
Walker, P.J., Kongsuwan, K., 1999. Deduced structural model for animal rhabdovirus glycoproteins. J. Gen. Virol. 80 (5), 1211–1220.
Walker, P.J., Byrne, K.A., Cybinski, D.H., Doolan, D.L., Wang, Y.H., 1991. Proteins of bovine ephemeral fever virus. J. Gen. Virol. 72 (1), 67–74.
Walker, P.J., Byrne, K.A., Riding, G.A., Cowley, J.A., Wang, Y., McWilliam, S., 1992. The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. Virology 191 (1), 49–61.
Walker, P.J., Wang, Y., Cowley, J.A., McWilliam, S.M., Prehaud, C.J., 1994. Structural and antigenic analysis of the nucleoprotein of bovine ephemeral fever rhabdovirus. J. Gen. Virol. 75 (8), 1889–1899.
Wang, Y., Walker, P.J., 1993. Adelaide river rhabdovirus expresses consecutive glycoprotein genes as polycistronic mRNAs: new evidence of gene duplication as an evolutionary process. Virology 195 (2), 719–731.
Wang, Y., McWilliam, S.M., Cowley, J.A., Walker, P.J., 1994. Complex genome organization in the GNS-L intergenic region of Adelaide River rhabdovirus. Virology 203 (1), 63–72.
Wertz, G.W., Moudy, R., Ball, L.A., 2002. Adding genes to the RNA genome of vesicular stomatitis virus: positional effects on stability of expression. J. Virol. 76 (15), 7642–7650.
Whelan, S.P., Barr, J.N., Wertz, G.W., 2004. Transcription and replication of nonsegmented negative-strand RNA viruses. Curr. Top Microbiol. Immunol. 283, 61–119.
Whitt, M.A., Zagouras, P., Crise, B., Rose, J.K., 1990. A fusion-defective mutant of the vesicular stomatitis virus glycoprotein. J. Virol. 64 (10), 4907–4913.
Wirblich, C., Tan, G.S., Papaneri, A., Godlewski, P.J., Orenstein, J.M., Harty, R.N., Schnell, M.J., 2008. PPEY motif within the rabies virus (RV) matrix protein is essential for efficient virion release and RV pathogenicity. J. Virol. 82 (19), 9730–9738.
Zhang, L., Ghosh, H.P., 1994. Characterization of the putative fusogenic domain in vesicular stomatitis virus glycoprotein G. J. Virol. 68 (4), 2186–2193.