Published in final edited form as:
Virus Res. 2011 September ; 160(1-2): 206–213. doi:10.1016/j.virusres.2011.06.021.

Genetic characterization of K13965, a strain of Oak Vale virus from Western Australia

Phenix-Lan Quan[a], **David T. Williams**[b], **Cheryl A. Johansen**[c], **Komal Jain**[a], **Alexandra Petrosov**[a], **Sinead M. Diviney**[b], **Alla Tashmukhamedova**[a], **Stephen K. Hutchison**[d], **Robert B. Tesh**[e], **John S. Mackenzie**[b,f], **Thomas Briese**[a,g,*], and **W. Ian Lipkin**[a,g,h]

[a]Center for Infection and Immunity, Columbia University, New York, NY, USA

[b]School of Biomedical Sciences, Curtin University, Perth, WA, Australia

[c]Discipline of Microbiology and Immunology, University of Western Australia, Nedlands, WA, Australia

[d]454 Roche Life Sciences, Branford, CT, USA

[e]Department of Pathology, University of Texas Medical Branch, Galveston, TX, USA

[f]Burnet Institute, Melbourne, Australia

[g]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

[h]Departments of Pathology and Neurology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Abstract

K13965, an uncharacterized virus, was isolated in 1993 from *Anopheles annulipes* mosquitoes collected in the Kimberley region of northern Western Australia. Here, we report its genomic sequence, identify it as a rhabdovirus, and characterize its phylogenetic relationships. The genome comprises a P′ (C) and SH protein similar to the recently characterized Tupaia and Durham viruses, and shows overlap between G and L genes. Comparison of K13965 genome sequence to other rhabdoviruses identified K13965 as a strain of the unclassified Australian Oak Vale rhabdovirus, whose complete genome sequence we also determined. Phylogenetic analysis of N and L sequences indicated genetic relationship to a recently proposed Sandjima virus clade, although the Oak Vale virus sequences form a branch separate from the African members of that group.

Keywords

rhabdovirus; Oak Vale virus; *Anopheles annulipes*; *Culex edwardsi*; virus phylogenetics; Australia; molecular characterization

*Corresponding author: Center for Infection and Immunity, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032, USA; Phone: 212-342-9031; Fax: 212-342-9044; thomas.briese@columbia.edu.

1. Introduction

Viruses with nonsegmented, single-stranded negative sense RNA genomes are subsumed in the order *Mononegavirales* that currently includes the four families *Bornaviridae*, *Filoviridae*, *Paramyxoviridae* and *Rhabdoviridae*. The family *Rhabdoviridae* comprises viruses of fish in the genus *Novirhabdovirus* and arthropod-borne viruses that infect plants in the genera *Nucleorhabdovirus* and *Cytorhabdovirus*. The remaining genera *Ephemerovirus*, *Lyssavirus* and *Vesiculovirus* include both arthropod-borne and non vector-borne pathogens of mammals, birds, reptiles and fish that include significant threats to agriculture and public health, such as vesicular stomatitis Indiana virus (VSIV) and rabies virus (RABV) (Kuzmin et al., 2009). In addition, more than 100 viruses are tentatively assigned to this family based on their characteristic bullet-shaped virion morphology, or via serologic cross-reactivity with one or more of the classified members of the family (Tordo et al., 2005).

Rhabdoviruses have genomes that range in approximate size from 11,000 to 16,000 nucleotides (nt). The genome organization of many classic rhabdoviruses, as represented by RABV, has been considered to be one of the simplest in the viral kingdom with only 5 genes: the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and polymerase (L) in the order 3′-N-P-M-G-L-5′ (Kuzmin et al., 2009). However, recent molecular analyses of novel rhabdoviruses indicated considerable genetic variability and more complex genomes, with up to 10 additional open reading frames (ORF) located within or interposed between the five core genes, many still without known function. In addition to overlapping ORFs characterized within the P gene of members of the *Vesiculovirus* genus (C and C′; (Kretzschmar et al., 1996; Peluso et al., 1996; Spiropoulou and Nichol, 1993), and in unassigned rhabdoviruses from vertebrates and dipterans (Allison et al., 2010; Gubala et al., 2010; Nishizawa et al., 1997; Quan et al., 2010; Schutze et al., 1999; Springfeld et al., 2005; Tao et al., 2008; Zhu et al., 2011), ORFs that overlap within N, M or G genes have also been identified (Gubala et al., 2008; Jayakar and Whitt, 2002; Tao et al., 2008). One to four ORFs located between the P and M genes are reported for plant-infecting rhabdoviruses (Dietzgen et al., 2006; Huang et al., 2003; Reed et al., 2005; Revill et al., 2005; Scholthof et al., 1994; Tanno et al., 2000), the unassigned Drosophila sigma viruses (Longdon et al., 2010; Teninges et al., 1993), and the *Culicoides*-transmitted Ngaingan (NGAV) and Wongabel viruses (WOGV) (Gubala et al., 2010; Gubala et al., 2008). In addition, an ORF located between the M and G genes has been identified in the unassigned NGAV, Durham (DURV), and Tupaia viruses (TUPV) (Allison et al., 2010; Gubala et al., 2010; Springfeld et al., 2005). All members of the genus *Novirhadovirus*, the nucleorhabdovirus rice yellow stunt virus (RYSV) and the unassigned Coastal Plains virus (CPV, Genbank accession GQ294473) have a unique ORF located between G and L genes, whereas some members of the genus *Ephemerovirus* contain ORFs potentially resulting from gene duplication at that junction (Huang et al., 2003; McWilliam et al., 1997; Morzunov et al., 1995; Wang et al., 1994). Additional genes located between the N and P genes have not yet been reported.

Arboviruses of medical importance have been monitored in northern Western Australia (WA) since the 1970s through serology and annual surveys of mosquitoes during the late wet season in major towns and communities across the Kimberley region (Broom et al., 2003; Liehne et al., 1981). During the course of those surveillance campaigns, a novel virus, designated K13965, was isolated. After K13965 failed to react with a panel of monoclonal antibodies raised to alphaviruses and flaviviruses common in WA (Broom et al., 1998), and did not yield amplification products using primers for the detection of common arboviruses, unbiased high-throughput pyrosequencing (UHTS) was employed for genetic characterization. Here we report the complete genome sequence of K13965, and its

identification as a rhabdovirus that represents a strain of Oak Vale virus (OVRV), for which we also report full genomic sequence.

2. Materials and Methods

2.1 Virus collection and culture

K13965 was isolated from a pool of female *Anopheles annulipes* s.l. mosquitoes collected in May 1993 from a trapping site located approximately 3 km north of Kununurra in the Kimberley region of WA (Fig. 1). Mosquitoes were collected in an annual survey of mosquitoes and arboviruses across northern WA by the University of Western Australia Arbovirus Surveillance and Research Laboratory (Broom et al., 2002; Johansen et al., 2009a). Mosquitoes were processed for virus isolation by inoculation of C6/36 cells, followed by serial passage of cell culture supernatant in PSEK and Vero cells (Johansen et al., 2000; Lindsay et al., 1993). Infected cultures were monitored by microscopic examination for cytopathic effect. Culture supernatants were harvested, serially passaged, and used for subsequent studies. Virus identification using infected cell monolayers was attempted by enzyme immunoassay using a panel of monoclonal antibodies to Australian arboviruses (Broom et al., 1998), and by reverse transcription-polymerase chain reaction (RT-PCR) assays on nucleic acid extracts using a panel of primer sets targeting alphaviruses, flaviviruses, orbiviruses and bunyaviruses (primer sequences available upon request).

OVRV strain CSIRO-1342 was obtained from the World Reference Center for Emerging Viruses and Arboviruses collection at the University of Texas Medical Branch. The isolation of OVRV strain CSIRO-1342 was reported from Peachester, Queensland, in 1981/82 from truck-trapped *Culex* sp. mosquitoes, and in 1984 from *Culex edwardsi* mosquito pools collected during an ephemeral fever outbreak (Calisher et al., 1989; Cybinski and Muller, 1990; Muller and Standfast, 1986) (Fig. 1).

2.2 Unbiased High-Throughput Sequencing (UHTS)

Culture supernatant from infected PSEK cells was clarified from cell debris by centrifugation and total RNA was extracted from the clarified supernatant for UHTS (Trizol-LS, Invitrogen, Carlsbad, CA, USA). RNA (0.5 μg) was DNase I-digested (DNA-free; Ambion, Austin, TX, USA) and reverse transcribed using Superscript II (Invitrogen) with random octamer primers linked to an arbitrary, defined 17-mer primer sequence. The cDNA was RNase H treated prior to random amplification by PCR using AmpliTaq (Applied Biosystems, Foster City, CA, USA) and a primer mix including the octamer-linked 17-mer sequence primer and the defined 17-mer sequence primer in a 1:9 ratio (Quan et al., 2007). Products >70 bp were purified (MinElute, Qiagen, Hilden, Germany) and ligated to linkers for sequencing on a GSL FLX Sequencer (454 Life Sciences, Branford, CT, USA) (Margulies et al., 2005). After trimming primer sequences and eliminating highly repetitive elements, sequences were clustered and assembled into contiguous fragments (contigs) for comparison by the Basic Local Alignment Search Tool (blast; (Altschul et al., 1990) to the Genbank database at nucleotide (nt) and deduced amino acid (aa) level, applying blastn and fastx algorithms.

2.3 Specific RT-PCR and quantitative real time RT-PCR

Multiple primer sets were designed based on sequences obtained through UHTS. The draft genome sequence was validated by sequencing overlapping PCR products that covered the entire genome. Genomic termini were characterized by RACE (Invitrogen). RNA was transcribed with Superscript II (Invitrogen) using random hexamer priming. PCR primers were applied at 0.2 μM concentration with 1 μL cDNA and Platinum Taq DNA polymerase

(Invitrogen). Products were purified (QIAquick PCR purification kit; Qiagen) and directly dideoxy-sequenced on both strands (Genewiz, South Plainfield, NJ, USA).

Primers and probe for quantitative real-rime PCR were selected within the L gene using Geneious software (Biomatters, Auckland, New Zealand): K13965-fwd 5′-TGGAGGAGACCATGACCAGCACA, K13965-rev 5′-GCTCAGACAGTTGGCTATGTTAGGAAG, K13965-probe 5′-FAM-AGCAAGAGGTATGCAAGAATGTGCA-TAMRA. A calibration standard was generated by cloning a 102 nt genomic fragment (pGEM-T Easy; Promega, Madison, WI, USA). Assays were performed in duplicate using a StepOnePlus Real-time PCR mix (Applied Biosystems), and a standard cycling profile with 45 cycles in a volume of 25 μL, containing random hexamer-primed cDNA, 300 nM primer (each) and 200 nM probe.

2.4 Phylogenetic analysis

Phylogenetic trees were constructed based on aa sequences of L (970 aa, N=49, or conserved block III, 156 aa, n=77) and N (325 aa, n=60). Sequences were aligned by the MUSCLE 3.6 software algorithm (Edgar, 2004) and manually adjusted by using MEGA 4.0 software (Tamura et al., 2007) after terminal regions with low alignment confidence were clipped. Phylogenetic trees were generated by applying the Bayesian method of the MrBayes software package (v 3.1.2) (Huelsenbeck and Ronquist, 2001), using a Whelan and Goldman (WAG) evolutionary model of aa replacement (Whelan and Goldman, 2001) with a gamma distribution of rate variation among sites. Chains were run for 10 million generations and sampled every 100th generation, applying a 25% burn-in rate at which point all parameters had converged.

2.5 Protein analysis

Similarity to protein families and prediction of functional protein domains were obtained through sequence analyses with PFAM (http://pfam.sanger.ac.uk/) and PROSITE (http://ca.expasy.org/prosite). Predictions of physico-chemical properties of deduced proteins of K13965 (molecular weight (MW), isoelectric point (pI), grand average of hydropathicity (GRAVY)) were generated by using the Protparam tool (http://expasy.ch/tools/). Phobius software (http://www.ebi.ac.uk/Tools/phobius/) was used to predict signal peptides and protein topology. Predictions of N-glycosylation and phosphorylation sites were derived with the respective algorithms of the Center for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/). Amino acid sequence identity and similarity were calculated by applying the Needleman algorithm with an EBLOSUM30 substitution matrix (gap open/extension penalties of 10/2 for aa alignments; EMBOSS (Rice et al., 2000)) and a Perl script to compile the results for all comparisons.

3. Results

3.1 K13965 genome organization

The complete K13965 genomic sequence was determined using infected cell culture supernatant. UHTS yielded approximately 112,800 sequence reads with a mean length of 245 nt. After primer trimming and assembly, 6 contigs ranging in size from 96 to 5,362 nt (14,363 sequence reads, mean length 353 nt) showed homology to rhabdovirus sequences in the NCBI non-redundant database (http://www.ncbi.nlm.nih.gov/Genbank). The assembled sequence covered full-length genome sequence, except for leader and trailer regions. Genomic termini were subsequently characterized by RACE. The generated draft sequence was confirmed by overlapping PCR across the whole genome (GenBank accession JF705877).

The negative-sense genome sequence of K13965 comprises 11,220 nt; the antigenome encodes at least eight ORFs (Fig. 2A). In addition to the five core rhabdoviral genes, K13965 includes an additional ORF (SH) located between the M and G genes, and two putative ORFs (P′, P″) in overlapping reading frames within the P gene to result in a genome organization 3′-Leader-N-P-(P′, P″)-M-SH-G-L-Trailer-5′. Putative transcription initiation (3′-CUGU) and termination (3′-GUACU$_7$) signals are recognized flanking the identified ORFs, resulting in six transcription units (Fig. 2A, B). Gene junctions show the usual separation of the upstream termination signal by one to two nt from the downstream initiation signal, except at the G-L junction. The intergenic regions consist of a dinucleotide (GA) at the N-P and SH-G junctions, whereas a single nucleotide (G or A) is present at the P-M and M-SH junctions (Fig. 2B). In contrast, the termination/polyadenylation signal of G is located 2 nt downstream of the L initiation sequence, resulting in overlap of the G and L genes without overlap of their ORFs.

A common feature of rhabdovirus genomes is partial complementary of their 3′- and 5′-termini (Whelan et al., 2004). The 3′-leader and the 5′-trailer sequences of K13965 comprise 39 and 51 nt, respectively; with eleven of the fifteen terminal nt being complementary (Fig. 2C).

3.2 Open reading frames

Features of the ORFs identified in the K13965 antigenome are summarized in Table 1. ORF-1 is compatible with nucleoproteins (N) of other rhabdoviruses by location, and contains a highly conserved aa motif S$_{289}$PYS that is proposed to be involved in RNA-binding in vertebrate rhabdoviruses (Crysler et al., 1990; Tordo et al., 1986). Alignment with other N sequences also indicates conserved aa identified as characteristic of rhabdoviruses that infect vertebrates (Kuzmin et al., 2006). ORF-II contains several potential phosphorylation sites, including 2 tyrosine, 24 serine, and 7 threonine phosphorylation sites as anticipated for a phosphoprotein (P), the cofactor of the active viral transcriptase/replicase complex. In addition, two non-overlapping ORFs in the +1 frame of the P gene may encode small acidic proteins designated P′ and P″ (Table 1, Figure 2). P′ is located close to the 5′-end of the P ORF, similar to C proteins identified in VSIV and vesicular stomatitis New Jersey virus (VSNJV) (Peluso et al., 1996; Spiropoulou and Nichol, 1993). Putative start codons for P′ and P″ are 65 nt and 758 nt downstream of the P start codon. Both putative proteins show no significant similarity to known proteins in the PFAM or PROSITE databases. ORF-III commonly codes for a matrix protein (M). The M analog of K13965 includes potential serine (6 sites) and threonine (1 site) phosphorylation sites, and no sequence similarity to other known proteins is detected. K13965 ORF-IV includes 2 hydrophobic aa clusters located at the N terminus and the center of the sequence, resulting in a grand average of hydropathicity of +0.595, analogous to the small hydrophobic (SH) proteins recently identified for TUPV and DURV (Allison et al., 2010; Springfeld et al., 2005). Phobius analysis indicated a signal peptide (aa 1-23) but no transmembrane anchor, suggesting that a soluble protein may be generated. Despite the limited primary sequence conservation amongst glycoproteins (G) of rhabdoviruses, general structural features including glycosylation sites, and cysteine residues are commonly conserved (Coll, 1995; Walker and Kongsuwan, 1999). Accordingly, ORF-V of K13965 shows features of type I glycoproteins, including an N-terminal signal peptide (aa 1-17), an ectodomain (aa 18-501), a transmembrane domain (aa 502-521) followed by a short cytoplasmic tail (aa 522-531), two potential N-glycosylation sites (N$_{378}$KTL and N$_{405}$GTT), and fourteen cysteine residues including 7 of the 12 cysteines conserved among animal rhabdoviruses (CI$_{40}$, CIII$_{81}$, CV$_{119}$, CVI$_{158}$, CVII$_{163}$, CXI$_{309}$, CXII$_{342}$; (Walker and Kongsuwan, 1999). The last and largest ORF-VI is predicted to translate into a 2,051 aa (231 kDa) L-polymerase. The L-polymerase of K13965 contains the conserved residues present in L-

polymerase of negative strand RNA viruses (Poch et al., 1990). In addition, all four highly conserved motifs A through D of block III (Poch et al., 1990; Poch et al., 1989) are conserved in L of K13965 (motif A (A_{577}NHMDYSKWNNHQR); motif B (A_{647}CWRGQAGGLEGLRQKGWTITSLLMI); motif C (T_{685}LAQGDNQIV); motif D (Y_{758}RGNLCNPKSKRY)).

3.3 Sequence comparison of K13965 with OVRV CSIRO-1342 and other rhabdoviruses

In an independent approach we also determined the complete genomic sequence of another unassigned Australian rhabdovirus, OVRV strain CSIRO-1342. As with K13965, culture supernatant of CSIRO-1342 was analyzed by UHTS, sequence gaps filled with specific PCR and RACE, and the draft sequence finally confirmed through re-sequencing by overlapping PCR (GenBank accession JF705876). The total RNA preparations from CSIRO-1342 and K13965 that were used for UHTS were subsequently evaluated by a real-time RT-PCR assay developed based on the consensus genome sequences. The quantitative OVRV assay indicated virus loads of 1.5x 10^5 copies/ng total RNA for CSIRO-1342, and 3.5 x 10^5 copies/ng for K13965.

Like K13965, the negative-sense genome of OVRV comprises 11,220 nt with a comparable genome organization. Furthermore, identical putative transcription initiation, termination and intergenic regions were identified for CSIRO-1342, except for a change at the M termination sequence (3′-G<u>C</u>ACU_7). As in K13965 the G and L genes overlap. The putative P′ ORFs of K13965 and CSIRO-1342 show high nt sequence conservation of 98.6%, and are identical at the aa level. However, the putative P″ start and stop codons are positioned differently in the same +1 frame, generating proteins of 67 aa for K13965, or 48 aa for CSIRO-1342. Comparison of OVRV CSIRO-1342 and K13965 genomes indicated 93 % nt identity. The P (92 %) and SH (93 %) ORFs show the least, and N (100 %) the highest aa sequence identity (Suppl. Table 1). Comparison of individual K13965 ORFs to homologous proteins of other representative rhabdoviruses for which complete sequence is available indicated that K13965 shares at most 36 % aa identity for L with TUPV and DURV; P and M show less then 17 % identity with the analyzed sequences, and identities for N ranged between 15 and 30 % (Suppl. Table 2).

3.4 Phylogenetic analyses

Phylogenetic trees constructed from partial L aa sequences of K13965, CSIRO-1342 and representative rhabdoviruses suggest that K13965 and CSIRO-1342 are closely related to, but distinct from, members of the recently proposed Sandjimba group (Fig. 3A) (Dacheux et al., 2010). Analysis of N aa sequences similarly grouped K13965 and CSIRO-1342 with the available Sandjimba group viruses KOLV and SJAV (Fig. 3B), but also showed a relationship to TUPV and DURV. Only the latter relationship could be confirmed by using nearly full-length L sequence due to a lack of sequence information for the other viruses (Fig. 3C).

4. Discussion

K13965, an uncharacterized Australian virus isolate, was genetically identified as a strain of OVRV by comparison of its genome to the full length genomic sequence that we obtained for the Australian OVRV CSIRO-1342, which matched the previously available partial sequence for CSIRO-1342 (GenBank Accession GQ294474). The K13965 genome, in addition to 5 core proteins typical for rhabdoviruses, encodes a small hydrophobic (SH) protein, and putative P′ (C) and P″ proteins that may be expressed from alternative ORFs contained within the P gene. Like sigma viruses in *Drosophila* spp. (Longdon et al., 2010;

Teninges et al., 1993), K13965, isolated from mosquitoes has overlap between the G and L gene without ORF overlap.

Within the order *Mononegavirales*, SH and C proteins were initially identified in paramyxoviruses (Bellini et al., 1985; Curran and Kolakofsky, 2008; Elango et al., 1989). Paramyxovirus C proteins have been shown to inhibit interferon signaling (Nagai and Kato, 2004), and SH proteins to inhibit TNF alpha production (Li et al., 2011) and apoptosis (Wilson et al., 2006). Protein products coded in analogous genome locations to paramyxovirus C proteins have been identified in the rhabdoviruses VSIV and VSNJV (Kretzschmar et al., 1996; Peluso et al., 1996; Spiropoulou and Nichol, 1993), and products comparable to SH proteins were identified in TUPV (Springfeld et al., 2005) and predicted for DURV (Allison et al., 2010). The functions of these proteins in rhabdoviruses are unknown although it has been hypothesized that C proteins may play a role in viral pathogenesis and enhance transcriptional activity (Peluso et al., 1996).

The putative P′ (C) of K13965 is larger than those reported for members of the genus *Vesiculovirus*, but smaller than P′ (C) reported for TUPV or DURV (Suppl. Table 3). The P′ (C) of K13965 is predicted to have an acidic pI whereas other rhabdoviral P′ (C) proteins, with the exception of that of DURV, have a basic pI >9. P′ (C) of K13965 shows no recognized sequence homology to other reported P′ (C), except for that of OVRV CSIRO-1342. In addition, a small P″ ORF located downstream of P′ was identified in K13965, resembling the P gene organization of NGAV, which is reported to also code for two putative non-overlapping small ORFs (Gubala et al., 2010). Although a P″ ORF is present also in CSIRO-1342, it is not conserved with respect to K13965; different start and stop codons predict a smaller protein for CSIRO-1342. This lack of conservation between the two isolates indicates a lower selective pressure on these sequences compared to the other coding sequences.

Both OVRV K13965 and CSIRO-1342 encode a putative hydrophobic SH protein of smaller size than those of TUPV and DURV (Suppl. Table 4) that has no sequence homology with these SHs. TUPV and DURV SHs are predicted to adopt a type I, or type II, transmembrane topology, respectively (Allison et al., 2010; Springfeld et al., 2005). In contrast, SH of OVRV is not predicted to be a transmembrane protein. Interestingly, in NGAV an ORF (U4; (Gubala et al., 2010) is present in the same genomic position as SH proteins that also likely generates a soluble but not hydrophobic protein (NC_013955; Phobius). No aa conservation was found in the N-terminal half of the ORF or the entire aa sequence. We also found no conservation of the leucine residues with respect to TUPV and DURV in the SH of K13965 or CSIRO-1342. Together with TUPV and DURV (isolated from mammal and bird, respectively), the mosquito-borne OVRV represents the third example of a rhabdovirus encoding a P′ (C) and SH protein.

The geographic region from where K13965-infected *Anopheles annulipes* s.l. mosquitoes were collected in 1993 is an enzootic focus for arboviruses, as the presence of the Ord River irrigation area enables mosquito species to breed continuously (Broom et al., 2002; Broom et al., 2003; Liehne et al., 1976; Mackenzie and Broom, 1999). The high level of rainfall during the 1992/1993 wet season was associated with an increase in arbovirus activity in that area (Broom et al., 1997). An isolate of the recently identified Stretch Lagoon orbivirus was also recovered from mosquitoes collected from the Ord River area in 1993 (D.T.W. and C.A.J., unpublished), suggesting that environmental conditions were suitable for heightened arbovirus activity at that time. OVRV CSIRO-1342 was isolated in the early 1980s from *Culex edwardsi* and *Culex* spp. mosquitoes from southeast Queensland (Calisher et al., 1989; Cybinski and Muller, 1990; Muller and Standfast, 1986). In addition, OVRV was also isolated in 1984 from *Aedes vigilax* mosquitoes collected in Darwin (Weir et al., 1997).

Taken together, OVRV strains have been isolated from disparate regions of tropical Australia, suggesting that this virus has been circulating widely for more than a decade.

An. annulipes s.l. is a species complex distributed across Australia, whilst *Ae. vigilax* has a coastal distribution and *Cx. edwardsi* is restricted to Queensland and the northern coast of New South Wales (Lee et al., 1989; Lee et al., 1987; Lee et al., 1984). Laboratory-based vector competence studies will be required to confirm whether mosquitoes of these species are vectors for OVRV. Blood meal studies suggest that *An. annulipes* s.l. and *Ae. vigilax* feed on a variety of vertebrate hosts, including birds, marsupials and cattle (Jansen et al., 2009; Johansen et al., 2009b; Lee et al., 1987; Lee et al., 1984), whereas *Cx. edwardsi* appear to prefer humans and birds (Lee et al., 1989). No serosurveys of antibodies to K13965 in humans or animals in the Ord River area have been conducted, and no antibodies to CSIRO-1342 were found in cattle sera at Peachester (Cybinski and Muller, 1990). OVRV likely cycles between mosquitoes and vertebrate host(s), possibly birds and/or pigs, as suggested by a reported detection of antibodies to OVRV in feral pigs (Bourhy et al., 2008).

Phylogenetic analysis of the block III L aa sequence indicates that OVRV is closely related to members of a recently proposed Sandjimba group (Dacheux et al., 2010) within the tentative Dimarhabdovirus supergroup (Bourhy et al., 2005) of the *Rhabdoviridae* family. Other members of this group were isolated from birds or mosquitoes from the Central African Republic. However, the Australian viruses K13965 and CSIRO-1342 form a branch separate from the African viruses, likely reflecting independent evolutionary pathways consistent with their geographic separation. In this regard it is of interest that the Sandjimba group forms a sister clade with the proposed Almpiwar group (Fig. 3), whose four members have been isolated from southern Queensland (Charleville virus strain Ch9824 from Charleville), from northern Queensland (Almpiwar virus and Charleville virus strain Ch9847 from Mitchell River), and from the Northern Territory (Humpty Doo virus from Beatrice Hill), similar geographic areas as Peachester, Darwin and the Kimberley region where OVRV and K13965 have been found (Fig. 1) (Bourhy et al., 2008). Thus, OVRV and Almpiwar group viruses have shared or overlapping geographic distributions. Analysis of the N aa sequence showed a similar relationship of OVRV to two viruses from the Sandjima group for which N sequence is available, and also indicated a relationship to TUPV and DURV that was corroborated by the long L sequence analysis; however, further analyses are hampered by the lack of long L and N sequences for other viruses, especially for Almpiwar group viruses. Additional genomic sequence information will be necessary to provide further insights into the evolutionary history and relationships of members of these proposed groups.

Arbovirus surveillance in northern Australia has yielded insights into the incidence, geographic range and ecology of major human flavi- and alphaviral pathogens through continued annual studies over the last 5 decades (Broom et al., 1998; Johansen et al., 2009a; Johansen et al., 2009b; Liehne et al., 1976). Our identification of Oak Vale rhabdovirus in uncharacterized isolates identified through those surveillance efforts further defines viral flora in that area and enhances our knowledge of the diversity and complexity of rhabdoviral genomes and their phylogenetic relationships.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Michael D.A. Lindsay (Arbovirus Surveillance and Research Laboratory, the University of Western Australia, Nedlands, Western Australia), Annette K. Broom (Arbovirus Surveillance and Research

Laboratory, the University of Western Australia, Nedlands, Western Australia), and Tony E. Wright (Mosquito Borne Disease Control Branch, Western Australian Department of Health, Claremont, Western Australia) for mosquito collection, and technical staff in the Arbovirus Surveillance and Research Laboratory for virus isolation. We also thank Craig Street for bioinformatic analyses. This work was supported by National Institutes of Health award AI57158 (Northeast Biodefense Center-Lipkin), contract HHSN27220100004OI/HHSN27200004/DO4, the United States Agency for International Development's (USAID) Emerging Pandemic Threats (EPT) Program, PREDICT project, under terms of Cooperative Agreement Number GHN-A-OO-09-00010-00, the United States Department of Defense and The Western Australian Department of Health by funding the Arbovirus Surveillance and Research Program.

References

Allison AB, Palacios G, Travassos da Rosa A, Popov VL, Lu L, Xiao SY, Detoy K, Briese T, Lipkin WI, Keel MK, Stallknecht DE, Bishop GR, Tesh RB. Characterization of Durham virus, a novel rhabdovirus that encodes both a C and SH protein. Virus Res. 2010; 155(1):112–22. [PubMed: 20863863]

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215(3):403–10. [PubMed: 2231712]

Bellini WJ, Englund G, Rozenblatt S, Arnheiter H, Richardson CD. Measles virus P gene codes for two proteins. J Virol. 1985; 53(3):908–19. [PubMed: 3882996]

Bourhy H, Cowley JA, Larrous F, Holmes EC, Walker PJ. Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene. J Gen Virol. 2005; 86(Pt 10):2849–58. [PubMed: 16186241]

Bourhy, H.; Gubala, AJ.; Weir, RP.; Boyle, DB. Animal rhabdoviruses. In: Mahy, BWJVR.; Van Regenmortel, MHV., editors. Encyclopedia of Virology. Vol. 3. 2008. p. 111-121.

Broom AK, Hall RA, Johansen CA, Oliveira N, Howard MA, Lindsay MD, Kay BH, Mackenzie JS. Identification of Australian arboviruses in inoculated cell cultures using monoclonal antibodies in ELISA. Pathology. 1998; 30(3):286–8. [PubMed: 9770194]

Broom AK, Lindsay M, van Heuzen B, Wright T, Mackenzie J, Smith D. Contrasting patterns of flavivirus activity in the Kimberley region of Western Australia, 1992–1996. Arbovirus Research in Australia. 1997; (7):25–30.

Broom AK, Lindsay MD, Harrington SA, Smith DW. Investigation of the southern limits of Murray Valley encephalitis activity in Western Australia during the 2000 wet season. Vector Borne Zoonotic Dis. 2002; 2(2):87–95. [PubMed: 12653302]

Broom AK, Lindsay MD, Wright AE, Smith DW, Mackenzie JS. Epizootic activity of Murray Valley encephalitis and Kunjin viruses in an aboriginal community in the southeast Kimberley region of Western Australia: results of mosquito fauna and virus isolation studies. Am J Trop Med Hyg. 2003; 69(3):277–83. [PubMed: 14628944]

Calisher CH, Karabatsos N, Zeller H, Digoutte JP, Tesh RB, Shope RE, Travassos da Rosa AP, St George TD. Antigenic relationships among rhabdoviruses from vertebrates and hematophagous arthropods. Intervirology. 1989; 30(5):241–57. [PubMed: 2676900]

Coll JM. The glycoprotein G of rhabdoviruses. Arch Virol. 1995; 140(5):827–51. [PubMed: 7605197]

Crysler JG, Lee P, Reinders M, Prevec L. The sequence of the nucleocapsid protein (N) gene of Piry virus: possible domains in the N protein of vesiculoviruses. J Gen Virol. 1990; 71 (Pt 9):2191–4. [PubMed: 1698928]

Curran J, Kolakofsky D. Nonsegmented negative-strand RNA virus RNA synthesis in vivo. Virology. 2008; 371(2):227–30. [PubMed: 18177685]

Cybinski DH, Muller MJ. Isolation of arboviruses from cattlee and insects at two sentinel sites in Queensland, Australia, 1979–85. Aust J Zool. 1990; 38:25–32.

Dacheux L, Berthet N, Dissard G, Holmes EC, Delmas O, Larrous F, Guigon G, Dickinson P, Faye O, Sall AA, Old IG, Kong K, Kennedy GC, Manuguerra JC, Cole ST, Caro V, Gessain A, Bourhy H. Application of broad-spectrum resequencing microarray for genotyping rhabdoviruses. J Virol. 2010; 84(18):9557–74. [PubMed: 20610710]

Dietzgen RG, Callaghan B, Wetzel T, Dale JL. Completion of the genome sequence of Lettuce necrotic yellows virus, type species of the genus *Cytorhabdovirus*. Virus Res. 2006; 118(1–2):16–22. [PubMed: 16313992]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5):1792–7. [PubMed: 15034147]

Elango N, Kovamees J, Varsanyi TM, Norrby E. mRNA sequence and deduced amino acid sequence of the mumps virus small hydrophobic protein gene. J Virol. 1989; 63(3):1413–5. [PubMed: 2915385]

Gubala A, Davis S, Weir R, Melville L, Cowled C, Walker P, Boyle D. Ngaingan virus, a macropod-associated rhabdovirus, contains a second glycoprotein gene and seven novel open reading frames. Virology. 2010; 399(1):98–108. [PubMed: 20089287]

Gubala AJ, Proll DF, Barnard RT, Cowled CJ, Crameri SG, Hyatt AD, Boyle DB. Genomic characterisation of Wongabel virus reveals novel genes within the *Rhabdoviridae*. Virology. 2008; 376(1):13–23. [PubMed: 18436275]

Huang Y, Zhao H, Luo Z, Chen X, Fang RX. Novel structure of the genome of Rice yellow stunt virus: identification of the gene 6-encoded virion protein. J Gen Virol. 2003; 84(Pt 8):2259–64. [PubMed: 12867659]

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17(8):754–5. [PubMed: 11524383]

Jansen CC, Webb CE, Graham GC, Craig SB, Zborowski P, Ritchie SA, Russell RC, van den Hurk AF. Blood sources of mosquitoes collected from urban and peri-urban environments in Eastern Australia with species-specific molecular analysis of avian blood meals. Am J Trop Med Hyg. 2009; 81(5):849–57. [PubMed: 19861621]

Jayakar HR, Whitt MA. Identification of two additional translation products from the matrix (M) gene that contribute to vesicular stomatitis virus cytopathology. J Virol. 2002; 76(16):8011–8. [PubMed: 12134006]

Johansen C, Broom A, Lindsay M, Avery V, Power S, Dixon G, Sturrock K, Maley F, McFall S, Geerlings K, Zammit C, Masters L, Bestall A, Smith D. Arbovirus and vector surveillance in Western Australia, 2004/05 to 2007/08. Arbovirus research in Australia. 2009a; 10:76–81.

Johansen CA, Power SL, Broom AK. Determination of mosquito (Diptera: Culicidae) bloodmeal sources in Western Australia: implications for arbovirus transmission. J Med Entomol. 2009b; 46(5):1167–75. [PubMed: 19769051]

Johansen CA, van den Hurk AF, Ritchie SA, Zborowski P, Nisbet DJ, Paru R, Bockarie MJ, Macdonald J, Drew AC, Khromykh TI, Mackenzie JS. Isolation of Japanese encephalitis virus from mosquitoes (Diptera: Culicidae) collected in the Western Province of Papua New Guinea, 1997–1998. Am J Trop Med Hyg. 2000; 62(5):631–8. [PubMed: 11289676]

Kretzschmar E, Peluso R, Schnell MJ, Whitt MA, Rose JK. Normal replication of vesicular stomatitis virus without C proteins. Virology. 1996; 216(2):309–16. [PubMed: 8607260]

Kuzmin IV, Hughes GJ, Rupprecht CE. Phylogenetic relationships of seven previously unclassified viruses within the family Rhabdoviridae using partial nucleoprotein gene sequences. J Gen Virol. 2006; 87(Pt 8):2323–31. [PubMed: 16847128]

Kuzmin IV, Novella IS, Dietzgen RG, Padhi A, Rupprecht CE. The rhabdoviruses: biodiversity, phylogenetics, and evolution. Infect Genet Evol. 2009; 9(4):541–53. [PubMed: 19460320]

Lee, DJ.; Hicks, MM.; Debenham, ML.; Griffiths, M.; Marks, EN.; Bryan, JH.; Russell, RC. The Culicidae of the Australasian Region, Entomology Monograph. Vol. 7. Australian Government Publishing Service; Canberra, Australia: 1989.

Lee, DJ.; Hicks, MM.; Griffiths, M.; Debenham, ML.; Bryan, JH.; Russell, RC.; Geary, M. The Culicidae of the Australasian Region, Entomology Monograph. Vol. 7. Australian Government Publishing Service; Canberra, Australia: 1987.

Lee, DJ.; Hicks, MM.; Griffiths, M.; Russell, RC.; Marks, EN. The Culicidae of the Australasian Region, Entomology Monograph. Vol. 7. Australian Government Publishing Service; Canberra, Australia: 1984.

Li Z, Xu J, Patel J, Fuentes S, Lin Y, Anderson D, Sakamoto K, Wang LF, He B. Function of the small hydrophobic protein of J paramyxovirus. J Virol. 2011; 85(1):32–42. [PubMed: 20980504]

Liehne PF, Anderson S, Stanley NF, Liehne CG, Wright AE, Chan KH, Leivers S, Britten DK, Hamilton NP. Isolation of Murray Valley encephalitis virus and other arboviruses in the Ord River Valley 1972–1976. Aust J Exp Biol Med Sci. 1981; 59(Pt 3):347–56. [PubMed: 6117274]

Liehne PF, Stanley NF, Alpers MP, Liehne CG. Ord River arboviruses--the study site and mosquitoes. Aust J Exp Biol Med Sci. 1976; 54(5):487–97. [PubMed: 14612]

Lindsay MD, Broom AK, Wright AE, Johansen CA, Mackenzie JS. Ross River virus isolations from mosquitoes in arid regions of Western Australia: implication of vertical transmission as a means of persistence of the virus. Am J Trop Med Hyg. 1993; 49(6):686–96. [PubMed: 8279636]

Longdon B, Obbard DJ, Jiggins FM. Sigma viruses from three species of Drosophila form a major new clade in the rhabdovirus phylogeny. Proc Biol Sci. 2010; 277(1678):35–44. [PubMed: 19812076]

Mackenzie, JS.; Broom, AK. Ord river irrigation area: the effect of dam construction and irrigation on the incidence of Murray Valley encephalitis virus. In: Kay, BH., editor. Water Resources - Health, Environment and Development. Spon; London: 1999. p. 108-122.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437(7057):376–80. [PubMed: 16056220]

McWilliam SM, Kongsuwan K, Cowley JA, Byrne KA, Walker PJ. Genome organization and transcription strategy in the complex GNS-L intergenic region of bovine ephemeral fever rhabdovirus. J Gen Virol. 1997; 78 (6):1309–17. [PubMed: 9191923]

Morzunov SP, Winton JR, Nichol ST. The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. Virus Res. 1995; 38(2–3):175–92. [PubMed: 8578857]

Muller, MJ.; Standfast, HA. Vectors of ephemeral fever group viruses. In: St George, TD.; Kay, BH.; Blok, J., editors. Arbovirus research in Australia: Proceedings of the Fourth Symposium. CSIRO/QMIR; Brisbane: 1986. p. 295-300.p. 295-298.

Nagai Y, Kato A. 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. 2004; 283:197–248. [PubMed: 15298171]

Nishizawa T, Kurath G, Winton JR. Sequence analysis and expression of the M1 and M2 matrix protein genes of hirame rhabdovirus (HIRRV). Diseases of Aquatic Organisms. 1997; 31:9–17.

Peluso RW, Richardson JC, Talon J, Lock M. 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. 1996; 218(2):335–42. [PubMed: 8610460]

Poch O, Blumberg BM, Bougueleret L, Tordo N. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J Gen Virol. 1990; 71 (Pt 5):1153–62. [PubMed: 2161049]

Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. Embo J. 1989; 8(12):3867–74. [PubMed: 2555175]

Quan PL, Junglen S, Tashmukhamedova A, Conlan S, Hutchison SK, Kurth A, Ellerbrok H, Egholm M, Briese T, Leendertz FH, Lipkin WI. Moussa virus: a new member of the *Rhabdoviridae* family isolated from Culex decens mosquitoes in Cote d'Ivoire. Virus Res. 2010; 147(1):17–24. [PubMed: 19804801]

Quan PL, Palacios G, Jabado OJ, Conlan S, Hirschberg DL, Pozo F, Jack PJ, Cisterna D, Renwick N, Hui J, Drysdale A, Amos-Ritchie R, Baumeister E, Savy V, Lager KM, Richt JA, Boyle DB, Garcia-Sastre A, Casas I, Perez-Brena P, Briese T, Lipkin WI. Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. J Clin Microbiol. 2007; 45(8):2359–64. [PubMed: 17553978]

Reed SE, Tsai CW, Willie KJ, Redinbaugh MG, Hogenhout SA. Shotgun sequencing of the negative-sense RNA genome of the rhabdovirus Maize mosaic virus. J Virol Methods. 2005; 129(1):91–6. [PubMed: 16005085]

Revill P, Trinh X, Dale J, Harding R. Taro vein chlorosis virus: characterization and variability of a new nucleorhabdovirus. J Gen Virol. 2005; 86(Pt 2):491–9. [PubMed: 15659770]

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000; 16(6):276–7. [PubMed: 10827456]

Scholthof KB, Hillman BI, Modrell B, Heaton LA, Jackson AO. Characterization and detection of sc4: a sixth gene encoded by sonchus yellow net virus. Virology. 1994; 204(1):279–88. [PubMed: 8091658]

Schutze H, Mundt E, Mettenleiter TC. Complete genomic sequence of viral hemorrhagic septicemia virus, a fish rhabdovirus. Virus Genes. 1999; 19(1):59–65. [PubMed: 10499451]

Spiropoulou CF, Nichol ST. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. J Virol. 1993; 67(6):3103–10. [PubMed: 8388490]

Springfeld C, Darai G, Cattaneo R. 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. 2005; 79(11):6781–90. [PubMed: 15890917]

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007; 24(8):1596–9. [PubMed: 17488738]

Tanno F, Nakatsu A, Toriyama S, Kojima M. Complete nucleotide sequence of Northern cereal mosaic virus and its genome organization. Arch Virol. 2000; 145(7):1373–84. [PubMed: 10963343]

Tao JJ, Zhou GZ, Gui JF, Zhang QY. Genomic sequence of mandarin fish rhabdovirus with an unusual small non-transcriptional ORF. Virus Res. 2008; 132(1–2):86–96. [PubMed: 18068257]

Teninges D, Bras F, Dezelee S. Genome organization of the sigma rhabdovirus: six genes and a gene overlap. Virology. 1993; 193(2):1018–23. [PubMed: 8384742]

Tordo N, Benmansour A, Calisher C, Dietzgen RG, Fang RX, Jackson AO, Kurath G, Nadin-Davis S, Tesh RB, Walker PJ. Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. 2005:623–644.

Tordo N, Poch O, Ermine A, Keith G, Rougeon F. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? Proc Natl Acad Sci U S A. 1986; 83(11):3914–8. [PubMed: 3459163]

Walker PJ, Kongsuwan K. Deduced structural model for animal rhabdovirus glycoproteins. J Gen Virol. 1999; 80 (Pt 5):1211–20. [PubMed: 10355768]

Wang Y, McWilliam SM, Cowley JA, Walker PJ. Complex genome organization in the GNS-L intergenic region of Adelaide River rhabdovirus. Virology. 1994; 203(1):63–72. [PubMed: 8030285]

Weir RP, Hyatt AD, CHC, Whelan PI. New records of arboviruses isolated from mosquitoes in the Northern Territory, 1982–1992. Arbovirus research in Australia. 1997; 7:311–321.

Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 2001; 18(5):691–9. [PubMed: 11319253]

Whelan SP, Barr JN, Wertz GW. Transcription and replication of nonsegmented negative-strand RNA viruses. Curr Top Microbiol Immunol. 2004; 283:61–119. [PubMed: 15298168]

Wilson RL, Fuentes SM, Wang P, Taddeo EC, Klatt A, Henderson AJ, He B. Function of small hydrophobic proteins of paramyxovirus. J Virol. 2006; 80(4):1700–9. [PubMed: 16439527]

Zhu RL, Lei XY, Ke F, Yuan XP, Zhang QY. Genome of turbot rhabdovirus exhibits unusual non-coding regions and an additional ORF that could be expressed in fish cell. Virus Res. 2011; 155(2):495–505. [PubMed: 21185339]

Highlights

Full genome sequence of Australian viruses K13965 and Oak Vale was determined by unbiased pyrosequencing. Genetic analysis identified K13965 as a rhabdovirus and as a strain of Oak Vale virus. Genome organization is unusual for P′(C)/P″ combined with a SH gene, and overlap of G and L genes. Phylogenetic analyses indicate relationships to the African Sandjimba clade and Australian Almpiwar clade.

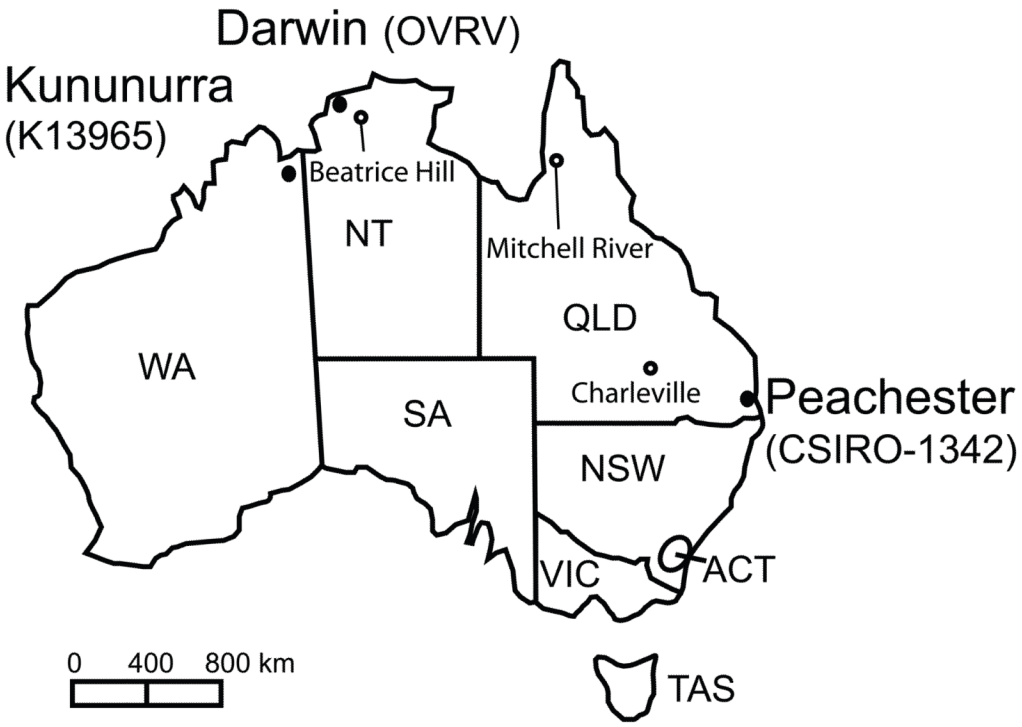
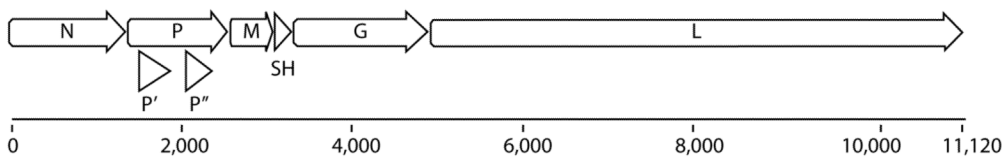


Figure 1.

Map of Australia by states and territories showing the geographic locations where OVRV K13965 and CSIRO-1342 were isolated. K13965 was isolated from *Anopheles annulipes* s.l. collected 3 km north of Kununurra. OVRV CSIRO-1342 was isolated from *Culex edwardsi* collected at Peachester. Another isolation of OVRV has been reported from *Aedes vigilax* collected at Darwin. ACT, Australian Capital Territory; NSW, New South Wales; NT, Northern Territory; QLD, Queensland; SA, South Australia; TAS, Tasmania; VIC, Victoria, WA, Western Australia.

A



B

Gene junction	Gene end	Intergenic site	Gene Start
Leader-N			CUGUGUGCAGUUGGUUC**UAC**
N-P	**AUC**CACGAAGUCUAUGUACUUUUUUUUGACUGUGUGCAGUGG**UAC**		
P-M	**ACU**CUUUGUACUUUUUUUUGCUGUGUGCAGUAG**UAC**		
M-SH	**AUU**GUACUUUUUUUUACUGUCUGGACUCGAGAAAGGUUC**UAC**		
SH-G	GU**ACU**UUUUUUUGACUGUGAGCAGUGGAAAUAG**UAC**		
G-L	**AUU**UGUCACUGUAGG**UAC**UUUUUUU		
L-Trailer	**AUU**GUGUACUUUUUUU		

C

```
Leader  UGCUCUUUUUGUUUUUACAAAAGUACAGCUGGUUUGUUAGCUGUGUGCAGUUGGUUCUAC
        |||| |||||| ||
Trailer ACGAAAAAAACCAAUCCAGAUGUUCUCCCCCCCAUCACCGGAUUCACCAACGUUUUUUUCAUGUGUUA
```

Figure 2.

K13965 genome organization. (A) Schematic organization of the K13965 genome. The bar represents the 11,120 nucleotide (nt) antigenome. The eight open arrows indicate the position of the open reading frames (ORFs) for putative N, P, P′, P″, M, SH, G, and L proteins. (B) Transcription initiation, intergenic and transcription termination/ polyadenylation sequences. The start and stop codon for each gene is in bold, the intergenic region is underlined. (C) Complementarity between the 3′ leader and the 5′ trailer sequences. Bases conserved among rhabdoviruses known to infect vertebrates are shown in bold. Start and stop codons are underlined.

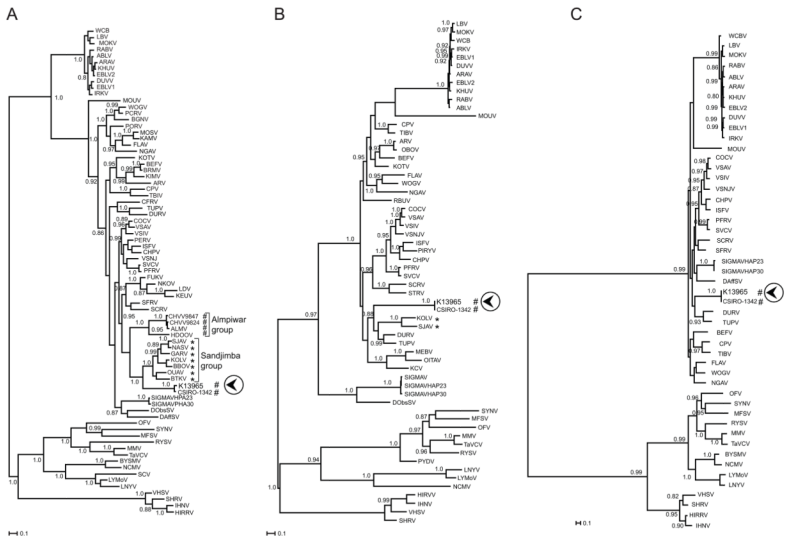


Figure 3.

Phylogenetic relationships of K13965 and CSIRO-1342 with representative rhabdoviruses derived from partial L (156 aa; **A**), N (325 aa; **B**) and full L amino acid sequence (970 aa; **C**). The geographic origin (*Africa, # Australia) is indicated for K13965, CSIRO-1342 and members of the Sandjimba and Almpiwar groups. K13965 is indicated by a circled arrowhead. Analysis was performed using a Bayesian method applying a WAG model of amino acid replacement with a gamma distribution of rate variation among sites for 10 million generations (with a 25% burn-in). Support for each node is provided by BPP (Bayesian posterior probability) (>0.8). Branch lengths are drawn to scale and the trees were mid-point rooted. Amino acid sequences used were: ABLV, Australian bat lyssavirus (L:NP_478343; N:AAD01267); ALMV, Almpiwar virus (L:AY854645); ARAV, Aravan virus (L:ABV03822; N:Q6X1D8); ARV, Adelaide River virus (L:AF234534; N:AAC54627); BBOV, Bimbo virus (L:GU816016); BEFV, bovine ephemeral fever virus (L:NP_065409; N:NP_065398); BGNV, Bangoran virus (L:GU816010); BRMV, Berrimah virus (L:AAZ43265); BTKV, Boteke virus (L:GU816014); BYSMV, barley yellow striate mosaic virus (L:ACT21686); CFRV, China fish rhabdovirus (L:AAX86686); CHPV, Chandipura virus (L:P13179; N:P11211); CHVV9824, Charleville virus (L:AY854644); CHVV9847, Charleville virus (L:AY854672); COCV, Cocal virus (L:ACB47438; N:ACB47434); CPV, coastal plains virus (L:ADG86364; N:ADG86356); DAffSV, Drosophila affinis sigma virus 10 (L:GQ410980); DobsSV, Drosophila obscura sigma virus 10A (N:GQ410979); DURV, Durham virus (L:ADB88761; N:ADB88758); DUVV, Duvenhage virus (L:ABZ81216; N:Q66453); EBLV1, european bat lyssavirus 1 (L:ABZ81181; N:AAX62875); EBLV2, european bat lyssavirus 2 (L:ABZ81191; N:YP_001285393); FLAV, Flanders virus (L:AAN73288; N:AAN73283); FUKV, Fukuoka virus (L:AAZ43279); GARV, Garba virus (L:GU816018); HDOOV, Humpty doo virus (L:AAZ43271); HIRRV, Hirame rhabdovirus (L:NP_919035; N:NP_919030); IHNV, infectious hematopoietic necrosis virus (L:CAA52076; N:NP042676); IRKV, Irkut virus (L:ABV03823; N:Q5VKP6); ISFV, Isfahan virus (L:Q5K2K3; N:Q5K2K7); KAMV, Kamese virus (L:GU816011); KCV, Kern Canyon virus (N:ABE69215); KEUV, Keuraliba virus (L:GU816021); KHUV, Khujand virus (L:ABV03824; N:Q6X1D4); KIMV, Kimberley virus (L:AAZ43266); KOLV, Kolongo virus (L:GU816020; N:ABE69214); KOTV, Kotonkan virus (L:AAZ43267; N:ABE69213); LBV, Lagos bat virus (L:ABZ81171; N:ABF56214); LDV, Le Dantec virus (L:AAZ43278); LYMoV, lettuce yellow mottle virus (L:YP_002308376; N:YP_00208371); LNYV, lettuce necrotic yellows virus (L:YP_425092; N:CAG34083); MEBV, mount Elgon bat virus (N:ABE69217);

MFSV, maize fine streak virus (L:YP_052849; N:YP_052843); MMV, maize mosaic virus (L:YP_052855; N:YP_052850); MOKV, Mokola virus (L:ABZ81211; N:YP_142350); MOSV, Mossuril virus (L:GU816012); MOUV, Moussa virus (L:ACZ81402; N:ACZ81403); NASV, Nasoule virus (L:GU816017); NCMV, northern cereal mosaic virus (L:NP_597914; N:NP_057954); NGAV, Ngaingan virus (L:YP_003518294; N:YP_003518280); NKOV, Nkolbisson virus (L:GU816022); OBOV, Obodhiang virus (N:ABE69212); OFV, orchid fleck virus (L:YP_001294929; N:BAH97109); OITAV, Oita rhabdovirus (N:BAD13431); OUAV, Ouango virus (L:GU816015); PCRV, Parry Creek virus (L:AAZ43275); PERV, Perinet virus (L:AAZ43280); PFRV, Pike fry rhabdovirus (L:ACP28002; N:ACP27998); PIRYV, Piry virus (N:P26037); PORV, Porton virus (L:GU816013); PYDV, potato yellow dwarf virus (N:ABW35154); RABV, Rabies virus (L:Q66T60; N:ACN15666); RBUV, Rochambeau virus (N:ABE69218); RYSV, rice yellow stunt virus (L:NP_620502; N:NP_620496); SCRV, Siniperca chuatsi rhabdovirus (L:YP_802942; N:YP_802937); SCV, strawberry crinkle virus (L:AAP03645); SFRV, Starry flounder rhabdovirus (L:AY450644); SHRV, snakehead rhabdovirus (L:NP_050585; N:NP_050580); SIGMAV, Drosophila melanogaster sigma virus (N:ACV67011); SIGMAVHAP23, Drosophila melanogaster sigma virus HAP23 (L:GQ375258; N:GQ375258); SIGMAVHAP30, Drosophila melanogaster sigma virus AP30 (L:YP_003126913; N:YP_003126908); SJAV, Sandjimba virus (L:GU816019; N:ABE69216); STRV, sea trout rhabdovirus (N:AAL35756); SVCV, spring viraemia of carp virus (L:Q91DR9; N:ABW24033); SYNV, Sonchus yellow net virus (L:NP_042286; N:NP_042281); TaVCV, Taro vein chlorosis virus (L:YP_224083; N:YP_224078); TIBV, Tibrogargan virus (L:ADG86355; N:ADG86347); TUPV, Tupaia virus (L:YP_238534; N:YP_238528); VHSV, viral hemorrhagic septicemia virus (L:CAB40833; N:P24378); VSAV, vesicular stomatitis Alagoas virus (L:ACB47443; N:ACB47439); VSIV, vesicular stomatitis Indiana virus (L:NP_041716; N:P11212); VSNJV, vesicular stomatitis New Jersey virus (L:P16379; N:P04881); WCBV, West Caucasian bat virus (L:ABV03821; N:Q5VKP2); WOGV, Wongabel virus (L:AAZ43276; N:YP_002333271).

Table 1

Characteristics of ORFs identified in the K13965 genome.

ORF	Gene length (nt)	ORF length (nt)	Protein length (aa)	Protein calculated MW (kDa)	pI	GRAVY
N	1390	1350	449	50.8	6.11	−0.302
P	1198	1170	389	41.4	5.58	−0.569
P′	-	276	91	10.4	6.10	−0.300
P″	-	204	67	7.6	6.83	−0.367
M	528	504	167	19.2	9.73	−0.140
SH	212	183	60	6.8	6.50	0.595
G	1637	1596	531	60.7	8.89	−0.296
L	6176	6156	2051	232	8.57	−0.203

pI: Isoelectric point; MW: molecular weight; GRAVY: Grand average of hydropathicity.

Table 2

Genome organization of representative rhabdoviruses.

Virus	Genus	Genome structure	Host/vector	Length (nt)	Genbank Accession no.
K13956	UC	**N-P**(P′, P″)-**M**-SH-**G-L**	mosquitoes	11120	JF705877
RABV	*Lyssavirus*	**N-P-M-G-L**	bat	11923	EU293116
VSIV	*Vesiculovirus*	**N-P**(C)-**M-G-L**	bovine	11155	AF473865
SIGMAV	UA	**N-P**-X-**M-G-L**	Drosophila melanogaster	12625	GQ375258
SCRV	UC	**N-P**(C)-**M** (Ms)-**G-L**	fish	11545	NC_008514
IHNV	*Novirhabdovirus*	**N-P-M-G**-NV-**L**	fish	11131	NC_001652
TUPV	Tentative *Vesiculovirus*	**N-P**(C)-**M**-SH-**G-L**	tree shrew	11440	NC_007020
RYSV	*Nucleorhabdovirus*	**N-P**-3-**M-G**-6-**L**	plant	14042	NC_003746
MFSV	*Nucleorhabdovirus*	**N-P**-3-4-**M-G-L**	plant	13782	NC_005974
TIBV	UA	**N-P-M**-U1-U2-**G(**G′**)**-U3-**L**	cattle, dipteran	13298	GQ294472
CPV	UC	**N-P**(P′)-**M**-U1-U2-**G**-U3-**L**	livestock	13203	GQ294473
WOGV	UC	**N**(U4)-**P**-U1-U2-U3-**M-G**(U5)-**L**	biting midges	13196	NC_011639
NCMV	*Cytorhabdovirus*	**N-P**-3-4-5-6-**M-G-L**	plant	13222	NC_002251
ARV	*Ephemerovirus*	**N-P-M-G**-G_{NS}-(α1/α2)-β-**L**	bovine	NA	U05987 U10363 L09206
BEFV	*Ephemerovirus*	**N-P(**P′**)-M-G**-G_{NS}-(α1/α2/α3)-(β/γ)-**L**	bovine	14900	NC_002526
NGAV	UA	**N-P** (P′1, P′2)-(U1/U2)-U3-**M**-U4-**G**-Gns-(U5/U6)-U7-**L**	biting midges	15764	NC_013955

UA: unassigned rhabdovirus; UC: unclassified virus (not found in the VIIIth report of the International Committee on taxonomy viruses). The 5 typical core genes N, P, M, G and L present in all rhabdoviruses are indicated in bold. ORFs translated from polycistronic RNA are indicated in parenthesis.