Virology 376 (2008) 13–23

Contents lists available at ScienceDirect

Virology

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

Genomic characterisation of Wongabel virus reveals novel genes within the *Rhabdoviridae*

Aneta J. Gubala [a,b,c,d,*], David F. Proll [c], Ross T. Barnard [b,d], Chris J. Cowled [a,b,d], Sandra G. Crameri [a], Alex D. Hyatt [a], David B. Boyle [a,d]

[a] CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, Victoria, Australia
[b] School of Molecular and Microbial Sciences, The 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

ARTICLE INFO

Article history:
Received 21 August 2007
Returned to author for revision
17 October 2007
Accepted 10 March 2008
Available online 24 April 2008

Keywords:
Wongabel virus
Rhabdovirus
Rhabdoviridae
Full genome sequence
Novel genes
Viroporin

ABSTRACT

Viruses belonging to the family *Rhabdoviridae* infect a variety of different hosts, including insects, vertebrates and plants. Currently, there are approximately 200 ICTV-recognised rhabdoviruses isolated around the world. However, the majority remain poorly characterised and only a fraction have been definitively assigned to genera. The genomic and transcriptional complexity displayed by several of the characterised rhabdoviruses indicates large diversity and complexity within this family. To enable an improved taxonomic understanding of this family, it is necessary to gain further information about the poorly characterised members of this family. Here we present the complete genome sequence and predicted transcription strategy of Wongabel virus (WONV), a previously uncharacterised rhabdovirus isolated from biting midges (*Culicoides austropalpalis*) collected in northern Queensland, Australia. The 13,196 nucleotide genome of WONV encodes five typical rhabdovirus genes N, P, M, G and L. In addition, the WONV genome contains three genes located between the P and M genes (U1, U2, U3) and two open reading frames overlapping with the N and G genes (U4, U5). These five additional genes and their putative protein products appear to be novel, and their functions are unknown. Predictive analysis of the U5 gene product revealed characteristics typical of viroporins, and indicated structural similarities with the alpha-1 protein (putative viroporin) of viruses in the genus *Ephemerovirus*. Phylogenetic analyses of the N and G proteins of WONV indicated closest similarity with the avian-associated Flanders virus; however, the genomes of these two viruses are significantly diverged. WONV displays a novel and unique genome structure that has not previously been described for any animal rhabdovirus.

Introduction

The *Rhabdoviridae*, along with the virus families *Paramyxoviridae*, *Filoviridae*, and *Bornaviridae*, belong to the order *Mononegavirales*. The *Rhabdoviridae* currently consists of six recognised genera, consisting of assigned and tentative species, and more than 120 yet unassigned isolates from around the world (International Committee on Taxonomy of Viruses, 2005). Members of the genera *Lyssavirus*, *Vesiculovirus* and *Ephemerovirus* have been isolated from a variety of different hosts, including mammals, fish and invertebrates. Several viruses that belong to these genera are known to cause significant disease in animals around the world, with potentially devastating impacts on livestock trade, wildlife, and humans. Members of the three other genera appear to be more host-specific. Members of the *Novirhabdovirus* genus are known to infect different species of fish, and members of the *Cytorhabdovirus* and *Nucleorhabdovirus* genera are arthropod-borne plant viruses.

Rhabdoviruses contain a negative-sense single stranded RNA genome that typically encodes five structural proteins; the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and polymerase (L), in the order N–P–M–G–L. However, many members of this family contain a more complex genome structure with a perplexing pattern of gene expression. Two viruses belonging to the genus *Ephemerovirus*, Bovine ephemeral fever virus (BEFV) and Adelaide River virus (ARV), have a large stretch of sequence between the G and L genes. This region contains a second glycoprotein gene (G_{NS}) and additional α, β and γ (BEFV) coding regions which are believed to have resulted from gene duplication (McWilliam et al., 1997; Walker et al., 1992; Wang et al., 1994; Wang and Walker, 1993). Viruses belonging to the genus *Novirhabdovirus* contain a single coding region between the G and L genes that encodes a non-virion (NV) protein, which has been shown to play an essential role in the formation of the cytopathic effect (CPE) in cell culture, and in pathogenicity for some species of fish (Basurco and

* Corresponding author. CSIRO Livestock Industries, Australian Animal Health Laboratory, Private Bag 24, Geelong Victoria 3220, Australia. Fax: +61 3 5227 5555.
E-mail address: ania.gubala@csiro.au (A.J. Gubala).

0042-6822/$ – see front matter.
doi:10.1016/j.virol.2008.03.004

Benmansour, 1995; Kurath and Leong, 1985; Thoulouze et al., 2004). The plant-infecting viruses belonging to the genera *Nucleorhabdovirus* and *Cytorhabdovirus* and the unclassified *Drosophila* Sigma virus (SIGMAV) all contain one to four additional genes between P and M, the function of which is unknown (Heaton et al., 1989; Landes-Devauchelle et al., 1995; Wetzel et al., 1994). The unassigned Tupaia rhabdovirus (TUPV) contains a novel small hydrophobic gene between the M and G genes (Springfeld et al., 2005) and the bird-infecting Flanders virus (FLAV) also appears to contain a complex genome structure. Genomic diversity is particularly prominent in members of the dipteran–mammalian "dimarhabdovirus supergroup", which includes viruses classified in the genera *Vesiculovirus* and *Ephemerovirus* and numerous unassigned viruses (Bourhy et al., 2005; Kuzmin et al., 2006). Currently available sequence data for these viruses suggests the need to establish new genera within this family, however, more viruses must be characterised to enable this. Although a large proportion of rhabdoviruses can cause diseases of significant importance, until recently only members of the genus *Lyssavirus* were recognised to cause deadly disease in humans. Recently, a large unexpected outbreak with a high fatality rate in India was attributed to Chandipura virus (CHPV), a member of the genus *Vesiculovirus* (Rao et al., 2004). This highlights the importance of characterising pathogens about which little is known: pathogens with an unidentified pathogenic potential.

In Australasia over the last five decades 18 different viruses have been isolated that have been tentatively assigned to the family *Rhabdoviridae* based on serological cross-reactivity or virion morphology. These viruses have been isolated from a diverse range of hosts, including mammals, mosquitos, flies and lizards. Serological cross-reactivity tests have indicated that some of these viruses share some relationships with each other but most are unique (Calisher et al., 1989; Karabatsos, 1985; Tesh et al., 1983). Little else is known about their evolution or relationships at the genetic level.

Wongabel virus (WONV) was isolated from biting midges (*Culicoides austropalpalis*) collected in 1979 at Wongabel on the Atherton Tablelands, northern Queensland, Australia by the CSIRO Long Pocket Laboratories, Brisbane, Australia (T. D. St George, personal communication). Limited information is available describing the isolation of this virus. *Culicoides austropalpalis* has a feeding preference for birds (Muller et al., 1981). Neutralizing antibodies to this virus were detected in 1.2% of sea bird sera collected off the Great Barrier Reef (Humphrey-Smith et al., 1991). No neutralizing antibodies to WONV have been detected in human sera collected from island residents within this region, and no link between WONV and disease has yet been established. WONV is recognised as a member of the family *Rhabdoviridae* based on its bullet-shaped morphology, however, it is not yet classified as a member of the *Rhabdoviridae* by the International Committee on Taxonomy of Viruses (ICTV) (International Committee on Taxonomy of Viruses, 2005). This report presents the characterisation of the entire WONV genome and its predicted transcription strategy, revealing three novel genes between the P and M genes and the presence of two additional open reading frames that overlap the 3′ ends of the N and G genes. The sequence of the full genome of this virus was obtained using a modified and improved PCR-select cDNA subtraction method. The function of the additional coding regions is unknown, however, their unexpected presence highlights the large diversity between these viruses and the need to gain a deeper understanding of the genomic organization of other uncharacterised rhabdoviruses.

Results and discussion

WONV morphology

Electron micrographs of WONV-infected BHK-BSR cells (Fig. 1) revealed bullet-shaped virions characteristic of viruses within the family *Rhabdoviridae*. Arrows indicated virions budding from the cell processes (Fig. 1A, B), at the plasma membrane (Fig. 1C) and from the endoplasmic reticula (Fig. 1D), which is an uncommon observation of viruses within this family. Virions were observed to be approximately 80–90×160–180 nm in dimensions. An intriguing observation was the common presence of elongated cell processes that contained virions lined up within, giving the appearance of a beaded string. Often, the particles within these cell processes appeared to be touching and the process appeared to be pinched in between, as highlighted by the rectangles in Fig. 1A and B. The reason for this occurrence is unclear; however, similar observations have been made for Tibrogargan virus (TIBV) (unpublished observations).

The complete genomic sequence of Wongabel virus

The PCR-select cDNA subtraction method is a reliable and efficient method for obtaining new sequence data for uncharacterised viruses (Bowden et al., 2001; Jack et al., 2005). This method has several considerable advantages over traditional cloning approaches for the sequencing of viruses. Firstly, no sequence information is required and no presumptions need to be made regarding the studied virus. Secondly, extensive virus purification is not required as total RNA from the infected tissue culture supernatant (TCSN) is used as the starting material. Furthermore, this method is very efficient for viruses with single stranded genomes, making the assembly of sequences faster and easier than for viruses with segmented genomes. With the aim of further improving the output of new sequence data generated by this method, we used alternate restriction enzymes, AluI and HaeIII, to digest cDNA, in addition to the enzyme RsaI provided in the kit. This generated three different restriction libraries of the WONV genome with overlapping fragments that allowed fast and simple the assembly of larger sequence contigs. This approach resulted in the generation of a large amount of sequence data with extensive coverage of the WONV genome. In total, 168 clones containing PCR products generated from the three individual cDNA subtractions were randomly selected. The inserts were sequenced and screened for virus-specific sequences by performing BLASTX similarity searches of the NCBI databases. A high proportion of the clones (124 clones; 74%) contained inserts that were rhabdoviral. The total amount of sequence data obtained from the three subtractions equated to 12,377 nt, representing ~95% coverage of the entire genome of WONV. This was in the form of five distinct contigs that were separated by only small regions of unknown sequence. Notably, sequence data was obtained to within as few as 23 nt from 3′ genome terminus and 25 nt from 5′ genome terminus, highlighting the power of this improved technique.

The sequence data obtained by the subtraction method was used to design 32 primer pairs which amplified overlapping products ranging in length from 500 to 800 nt. These PCR primers were used to amplify the entire WONV genome, excluding the terminal regions. Following the sequencing of these PCR products, a single consensus sequence of 13,148 nt, representing the majority of the WONV genome, was created. Finally, the sequences of the genome termini were obtained by a modified RACE method to complete the sequence of the entire 13,196-nt genome of WONV. The average quality scores of the sequence data determined by the SeqMan program were very good. Every nucleotide of the WONV genome was sequenced with a minimum four-fold coverage in each direction, including at least one-fold coverage in each direction directly from PCR products as template. The majority of nucleotides had more than 10-fold coverage. Collectively, this provided high confidence in the final sequence.

The organization of the 13,196-nt genome of WONV is depicted in Fig. 2. The genome contains the five typical rhabdovirus genes; N, P, M, G and L. The genome also contains three additional genes between the P and M genes, and the G and N genes contain long 3′ untranslated regions (UTRs) that appear to contain additional ORFs that overlap with the 3′ end of the preceding ORF. The presence of these five additional genes and/or ORFs suggests that WONV has a novel and unique genome organisation among animal rhabdoviruses.