Short Communication

Complete genome analysis identifies Tvarminne avian virus as a candidate new species within the genus Orthoreovirus

Eszter Dandár,1 Eili Huhtamo,2 Szilvia L. Farkas,1 Miklós Oldai,3,4 Ferenc Jakab,3,4 Olli Vapalahti2,5,6 and Krisztián Bányaí1

1Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary
2Department of Virology, Haartman Institute, Faculty of Medicine, University of Helsinki, Helsinki, Finland
3Institute of Biology, Faculty of Sciences, University of Pécs, Pécs, Hungary
4János Szentagotai Research Center, University of Pécs, Pécs, Hungary
5Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland
6Department of Virology and Immunology, HUSLAB, Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Orthoreoviruses have been associated with a variety of diseases in domesticated poultry and wild-living birds. In 2002, a reovirus strain named Tvarminne avian virus (TVAV), was identified in Finland in a crow showing neurological disorders. The objective of this study was the molecular characterization of this novel reovirus strain. Genome sequencing was performed by combining semiconductor sequencing and traditional capillary sequencing. Sequence and phylogenetic analyses showed that TVAV shares low nucleotide sequence identity with other reoviruses (range for each gene, 31–72 %) including strains belonging to the species Avian orthoreovirus. The most closely related reovirus strain was an isolate identified in Steller sea lion. Our data indicate that TVAV is a divergent reovirus of avian origin that may be the first representative of a distinct virus species within the genus Orthoreovirus.

Viruses of the family Reoviridae are classified into two subfamilies, Spinareovirinae and Sedoreovirinae and 15 genera. Orthoreoviruses are listed in the subfamily Spinareovirinae and currently divided into five accepted species: Mammalian orthoreovirus (MRV); Avian orthoreovirus (ARV); Nelson Bay orthoreovirus (NBV); Baboon orthoreovirus (BRV); Reptilian orthoreovirus (RRV) (Attoui et al., 2011). Characteristic to all orthoreoviruses is their linear dsRNA genome that comprises 10 segments classified into three groups, large (L1–L3), medium (M1–M3) and small (S1–S4), based on their size and mobility in agarose gels. The coding regions of each segment are flanked by the 5′- and 3′-end UTRs. With the exception of the S1, or in some strains the S4, all orthoreovirus genome segments are monocistronic. The number of structural and non-structural genes shows some variation across orthoreoviruses, but in general, seven or eight genes have a role in capsid formation and at least three genes encode non-structural proteins (Benavente & Martinez-Costas, 2007).

Modern species demarcation for orthoreoviruses is based on several criteria including: (i) nucleotide and amino acid sequence identity of the homologous genome segments (in most cases >75 % nucleotide sequence identity within species versus <60 % between species) and proteins (for the more conservative core proteins >85 % within species and <65 % identity between species, for the outer and more divergent capsid proteins >55 % versus <35 %, respectively); (ii) reassortment between members of the same but not those from other species; (iii) the RNA fingerprint; (iv) structure of the polycistronic genome segment which is either bi- or tricistronic; and (v) host specificity and observed clinical signs (Duncan, 1999; Duncan et al., 2004; Jones, 2008; Schiff et al., 2007).

Phylogenetic analyses of cognate genes have revealed notable clustering of several representative orthoreovirus species, indicating that virus diversification and speciation...
have occurred at various time points during the evolution of this genus. Molecular phylogenetic studies based on presently available data indicate that MRVs represent a major divergent clade; NBVs and ARVs form another clade (Duncan, 1999). This clade includes some newly discovered and analysed strains (psittacine reoviruses and Steller sea

<table>
<thead>
<tr>
<th>Genome segment</th>
<th>Size (bp)</th>
<th>Length (bp) of</th>
<th>Sequence at the termini</th>
<th>Protein size (bp)</th>
<th>Encoded protein*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5’-end UTR</td>
<td>ORF</td>
<td>3’-end UTR</td>
<td>5’ end</td>
</tr>
<tr>
<td>L1</td>
<td>3993</td>
<td>25</td>
<td>3888</td>
<td>80</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>L2</td>
<td>3901</td>
<td>12</td>
<td>3852</td>
<td>36</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>L3</td>
<td>3829</td>
<td>13</td>
<td>3780</td>
<td>38</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>M1</td>
<td>2288</td>
<td>13</td>
<td>2214</td>
<td>61</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>M2</td>
<td>2154</td>
<td>29</td>
<td>2028</td>
<td>97</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>M3</td>
<td>2026</td>
<td>25</td>
<td>1920</td>
<td>81</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>S1</td>
<td>1659</td>
<td>22</td>
<td>297</td>
<td>81</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>S2</td>
<td>1325</td>
<td>15</td>
<td>1251</td>
<td>59</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>S3</td>
<td>1202</td>
<td>29</td>
<td>1104</td>
<td>69</td>
<td>GCCUUUC</td>
</tr>
<tr>
<td>S4</td>
<td>1201</td>
<td>22</td>
<td>1107</td>
<td>72</td>
<td>GCCUUUC</td>
</tr>
</tbody>
</table>

*Protein nomenclature was adopted from Nibert & Duncan (2013). RdRp, RNA-dependent RNA polymerase.

Fig. 1. Comparative diagram based on the percentage nucleotide sequence identities of the different genome segments between the TVAV and the representative orthoreovirus strains of species MRV, ARV, NBV, RRV, BRV, BroV and SSRV. The grey area indicates the species demarcation cut-off values.
lion reovirus; SSRV), whereas RRV and BRV form a third cluster together with Broome virus (BroV) suggesting their common evolutionary history (Bányai et al., 2013; Palacios et al., 2011; Thalmann et al., 2010). In fact, the established similarity based species demarcation criteria are invalid in some pairwise comparisons. For example, NBV and ARV
Phasianus colchicus reovirus was detected in a pheasant (Chappel et al., 2000; Jones, 2000). In wild-living species, reovirus was detected in a pheasant (Phasianus colchicus) with diarrhoea (Mutlu et al., 1998), it was associated with wasting in American woodcock (Scolopax minor) (Docherty et al., 1994) and gastrointestinal disease in young Virginia quails (Colinus virginianus) (Ritter et al., 1986), and caused death of common eider ducklings (Somateria mollissima) (Hollmén et al., 2002). A wild hooded crow (Corvus corone cornix) showing CNS signs was found in Finland in 2002 (Huhhtamo et al., 2007). From the brain of the bird a reovirus, Tvarmminne avian virus (TVAV), with syncytium forming capacity was isolated. Based on sequence and phylogenetic analysis of partial gene sequences encoding the σC and μNS proteins, TVAV was distinct from the previously described avian orthoreoviruses and formed a separate clade in the genus Orthoreovirus. These data raised uncertainty in the taxonomy of TVAV which has remained unresolved since its first description. The objective of this study was to clarify the genetic relatedness of TVAV to ARVs and other members of the genus. In light of this objective the whole genome share up to 68% nucleotide and 75% amino acid similarity in the λA gene.

At present, all avian origin orthoreoviruses are classified into the ARV species (Day, 2009). ARVs with some exceptions (e.g. several goose and Muscovy duck reovirus strains) possess a fusion associated small transmembrane (FAST) protein (also referred to as p10 in ARVs) enabling them to generate a typical cytopathic effect, giant cell formation in cell cultures. ARVs have been detected in economically important, exotic and wild-living fowl species. For example, reovirus infections in chicken usually manifest as tenosynovitis, runting-stunting or malabsorption syndrome, growth retardation, occasionally central nervous system (CNS) signs, respiratory diseases, hepatitis, hydropericardium, "blue-wing disease" and/or sudden death (Chappel et al., 2000; Jones, 2000).
sequence was determined and each genome segment was compared to cognate genes of strains within the genus.

To obtain sufficient genomic RNA for whole genome sequencing, TVAV was propagated on baby hamster kidney (BHK-21) cells using a 75 cm² flask and was harvested by freezing and thawing on day 4 post-infection. Virus concentration, RNA extraction, random primed reverse transcription-PCR, library construction, template enrichment, and semiconductor sequencing on Ion Torrent PGM equipment was carried out as described previously for a bush viper reovirus (Bányai et al., 2014). Next-generation sequencing yielded a draft genome of TVAV at >50 x mean coverage. Subsequently, the authentic terminal sequences were determined by an RNA ligation method as described elsewhere in detail (Bányai et al., 2011; Lambden et al., 1992). In addition, genomic regions that were not fully resolved by massive parallel sequencing were sequenced by the Sanger sequencing method on ABI Prism 310 equipment using gene specific primers. The genome sequence of TVAV was deposited in GenBank under the following accession numbers: KF692089–KF692098.

The software MIRA was used to clean and assemble next-generation sequencing data (Chevreux et al., 1999). Contigs were manually adjusted and edited with Sanger sequencing reads in the GeneDoc software (Nicholas et al., 1997). ORFs were identified by ORFinder (http://www.ncbi.nlm.nih.gov/orf/orf.html). BLASTN, BLASTX and BLASTP algorithms were used to identify homologous genes among sequences deposited in GenBank (Altschul et al., 1990). The Muscle algorithm within the TranslatorX online platform was used to obtain codon based multiple alignments (Abascal et al., 2010). Determination of nucleotide and amino acid sequence similarity values and phylogenetic analysis was performed using the MEGA5 package (Tamura et al., 2011). Best fit substitution models were selected based on the Bayesian information criterion. Neighbour-joining and maximum-likelihood (ML) trees were generated and tree topologies were validated by bootstrap analysis as implemented in MEGA5 (Tamura et al., 2011).

In brief, the genome of TVAV measured 23,578 bp; individual genome segment sizes ranged from 3993 bp (L1) to 1201 (S4). BLAST search identified all major orthoreovirus gene homologues (Table 1). The polycistronic S1 genome segment encoded three putative proteins homologous to chicken ARV proteins: p10, p17 and σC, respectively. The lengths of UTRs ranged from 12 to 29 bp at the 5’ ends and 32 to 96 bp at the 3’ ends. Each genome segment started and ended with short, highly conserved sequences, of which the sequence at the 5’ end (GCCUUUC) was different from the sequences determined for other orthoreovirus species, whereas the sequence at the 3’ end (A/UUCAUC) was a typical orthoreovirus genome segment termination motif (Benavente & Martinez-Costas, 2007). The sequence similarity values of individual genes and the proteins they encode showed remarkable divergence from type strains of all five orthoreovirus species and the additional unclassified orthoreovirus strains (Fig. 1). The greatest sequence similarity values were seen with SSRV (up to 72% nucleotide and 83% amino acid similarity for the gene encoding μB). The greatest sequence identity values with a reference ARV strain (AVS-B) were observed for the λA encoding gene (71% nucleotide, 84% amino acid). Six out of 10 genes shared <60% nt identity (the cut-off value for species demarcation) with cognate genes of representative strains of the established orthoreovirus species; only three inner core genes (λA, JB, σA) and an outer-shell protein coding gene, μB, showed >60% but <75% nucleotide identity.

Phylogenetic analysis of all major genes revealed that they cluster with the ARV-like strain, SSRV, and in most cases this relationship was supported by high bootstrap values (Fig. 2). However, the phylogeny of sigma-class genes revealed that SSRV is more similar to psittacine orthoreoviruses, whereas TVAV is only distantly related to the cluster of these divergent orthoreovirus strains. Only the gene encoding σA of TVAV was phylogenetically more similar to typical ARVs. This finding suggests that reassortment with a heterologous ARV-like orthoreovirus could have played a role in the origin of TVAV, even though the statistical support for this topology was not very strong. Collectively, it was apparent that TVAV clusters with reference strains of ARVs, NBVs and ARV-like reoviruses (such as those detected in parrots and Steller sea lion). This topology reinforced the existence of a large monophyletic group of these orthoreoviruses (Chappell et al., 2000), where divergent strains originating from various hosts, causing different clinical diseases in their respective hosts, and sharing various extents of sequence similarities in genes encoding proteins of the core and the outer-shell virion, are located.

TVAV is an avian origin orthoreovirus which was isolated from a crow with neurological disorders. Reovirus associated neurotropism is not unusual: type 1 and 3 MRVs cause CNS infection in mice (Phillips et al., 1970), BRV causes meningoencephalomyelitis in baboons (Duncan et al., 1995), and some chicken ARV strains have been identified in chickens with CNS diseases (Dandár et al., 2013). Thus, the role of TVAV in CNS disease in corvid birds seems possible based on literature data; however, formal demonstration of such disease association would require experimental data, or, alternatively the collection of additional epidemiological data.

As the sequence analysis results did not exclusively show TVAV to be associated with bird hosted orthoreoviruses, it seems possible that the crow was merely an incidental host for TVAV. For the other cluster of ARV-like reovirus strains, which include psittacine reoviruses and SSRV, both avian and mammalian hosts are recognized. In addition, NBVs were detected in Megachirotekta bats and humans, causing acute respiratory illness in the latter (Chua et al., 2007, 2008, 2011; Pritchard et al., 2006). The fact that TVAV could be isolated and grown on a hamster origin tissue culture in vitro (Huhtamo et al., 2007) and the preliminary experimental evidence that TVAV killed embryonic chicken when it was passed in hen’s egg (E.
Dandaır, unpublished data) might also indicate a limited host fidelity of TVAV. All these findings imply that the crow-isolated TVAV is clearly not a typical avian orthoreovirus isolate, and it seems possible that several orthoreovirus strains that cluster with the former ARV-NBV clade may have unusual and perhaps an extended host range.

Inevitably, sequence similarity values are key elements in modern virus taxonomy and intra-species classification schemes. Our molecular analyses demonstrated that TVAV shares only low sequence similarity to representative strains of any other hitherto characterized orthoreoviruses, except some conserved virion protein coding genes, which were somewhat more similar to cognate genes of typical ARVs and NBVs (Figs 1–3). However, in these cases the similarity values fell in the grey zone of the species demarcation cut-off values. Of interest, the cut-off values of the same series of genes also fell in the grey zone when NBVs were compared to ARV strains (data not shown), even if they are classified into separate virus species. The question of whether TVAV and SSRV should be classified into a single common or separate orthoreovirus species remains open. The phylogenetic analyses which demonstrate their common evolutionary roots are neither definitive nor conclusive, given that similar clustering of NBVs and ARVs was seen when only limited sequence data were available. In light of current species demarcation similarity scores as well as the differences in the 5′-end termination sequences, the host species and the clinical signs clearly indicate that TVAV and SSRV (plus psittacine reoviruses) are only distant relatives of each other.

In conclusion, the identification of novel, genetically divergent orthoreoviruses, such as SSRV, psittacine reoviruses and TVAV, raise some virus taxonomic issues. In this study, we supplemented a previous report on the occurrence, clinical features and partial virus characterization data of TVAV by determining the whole virus genome sequence. We demonstrated that TVAV shares only low nucleotide and amino acid sequence similarity with other orthoreovirus strains, and thus could be considered as a representative of a new species within the genus Orthoreovirus. Accumulation of more data will help to define the host relationship of TVAV; however, a provisional name *Corvid orthoreovirus* (CRV) could be given implying the currently known host species of this divergent virus.

**Acknowledgements**

The study was supported by the Hungarian Scientific Research Fund (OTKA) (K108727) and a grant from the Hospital District of Helsinki and Uusimaa (EVO/TYH2011305). Additional support was given by the Momentum Programme (K.B.), the Emil Aaltonen Foundation (E.H.) and the Hungarian National Excellence Programme (TÁMOP 4.2.4. A/2-11-1-2012-0001) (F.J.).

**Fig. 3.** Whole-genome-based mVISTA nucleotide sequence alignments comparing TVAV with representative orthoreovirus strains of species MRV, ARV, NBV, RRV, BRV, BroV and SSRV. The size of each genome segment has been adjusted. The height of the contoured area at any sampling point is proportional to the genetic relatedness. The grey regions indicate identity >75%. S1 refers to the α1/αC protein coding ORF; this genomic region cannot be found in BRV and BroV. The scale bar shows the approximate length of the concatenated genomes.