

Euprosterina elaeasa virus genome sequence and evolution of the *Tetraviridae* family: Emergence of bipartite genomes and conservation of the VPg signal with the dsRNA *Birnaviridae* family[☆]

Jean-Louis Zeddou^{a,b}, Karl H.J. Gordon^c, Chris Lauber^d, Cristiano A. Felipe Alves^e, Brian T. Luke^f, Terry N. Hanzlik^c, Vernon K. Ward^e, Alexander E. Gorbalenya^{d,*}

^a IRD, UR 072, Laboratoire Evolution, Génomes et Spéciation, UPR 9034, CNRS, 91198 Gif-sur-Yvette Cedex, France

^b Université Paris-Sud 11, 91405 Orsay Cedex, France

^c CSIRO Entomology, GPO Box 1700, Canberra ACT 2601, Australia

^d Molecular Virology Laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Postzone E4-P, P.O. Box 9600, NL-2300 RC Leiden, The Netherlands

^e Department of Microbiology and Immunology, School of Medical Sciences, University of Otago, PO Box 56, Dunedin, 9054, New Zealand

^f Advanced Biomedical Computing Center, Science Applications International Corporation/National Cancer Institute, P.O. Box B, Frederick, MD 21702-1201, USA

ARTICLE INFO

Article history:

Received 27 August 2009

Returned to author for revision

8 October 2009

Accepted 28 October 2009

Available online 1 December 2009

Keywords:

Tetraviridae

Birnaviridae

Insect virus

Genome sequence

Permuted RNA-dependent RNA polymerase

VPg

Jelly-roll-based capsid

Phylogeny

Recombination

Taxonomy

ABSTRACT

The *Tetraviridae* is a family of non-enveloped positive-stranded RNA insect viruses that is defined by the T=4 symmetry of virions. We report the complete *Euprosterina elaeasa* virus (EeV) genome sequence of 5698 nt with no poly(A) tail and two overlapping open reading frames, encoding the replicase and capsid precursor, with ~67% amino acid identity to *Thosea asigna* virus (TaV). The N-terminally positioned 17 kDa protein is released from the capsid precursor by a NPGP motif. EeV has 40 nm non-enveloped isometric particles composed of 58 and 7 kDa proteins. The 3'-end of TaV/EeV is predicted to form a conserved pseudoknot. Replicases of TaV and EeV include a newly delineated VPg signal mediating the protein priming of RNA synthesis in dsRNA *Birnaviridae*. Results of rooted phylogenetic analysis of replicase and capsid proteins are presented to implicate recombination between monopartite tetraviruses, involving autonomization of a sgRNA, in the emergence of bipartite tetraviruses. They are also used to revise the *Tetraviridae* taxonomy.

© 2009 Elsevier Inc. All rights reserved.

Introduction

The *Tetraviridae* is a lepidopteran-restricted family of viruses with positive-sense ssRNA [ssRNA+] genomes (Entwistle, 1987; Moore, 1991; Hanzlik and Gordon, 1997; Gordon and Hanzlik, 1998; Hanzlik et al., 1999; Christian et al., 2001; Gordon and Waterhouse, 2006). Tetravirus particles are 35–41 nm in diameter and composed of 240 copies of two proteins of approximately 60 kDa (L) and 8 kDa (S). The structures of *Nudaurelia capensis* β virus (N β V) and *Nudaurelia capensis* ω virus (N ω V) have been solved (Finch et al., 1974; Olson et al., 1990; Johnson et al., 1994; Munshi et al., 1996, 1998) and reveal T=4 icosahedral capsid architecture that

distinguishes the family from other non-enveloped viruses. Eleven viruses are recognized in the family (Hanzlik et al., 2005), and N β V (Gordon et al., 1999), N ω V (Agrawal and Johnson, 1992) and *Helicoverpa armigera* stunt virus (HaSV) (Gordon et al., 1995; Hanzlik et al., 1995) are considered prototypic tetraviruses.

The *Tetraviridae* contains the *Betatetravirus* and the *Omegetetravirus* genera (Hanzlik et al., 2005). *Betatetraviruses*, such as N β V, have monopartite genomes; the 6.5 kb genomic RNA encodes the replicase and capsid precursor genes. Virions also encapsidate a 2.5 kb subgenomic RNA (sgRNA), identical to the 3'-region of the genomic RNA, that directs synthesis of the capsid precursor (see Fig. 1) (Gordon et al., 1999; Pringle et al., 1999, 2003). *Betatetravirus* capsids display three distinct pits and each face is separated by a deep groove (Olson et al., 1990; Pringle et al., 1999). *Omegetetraviruses*, such as N ω V and HaSV, have bipartite ssRNA+ genomes of approximately 5.2 and 2.5 kb encoding the replicase and capsid precursor genes, respectively. *Omegetetravirus* capsids are more

[☆] The GenBank accession number of the EeV genome sequence described in this paper is AF461742.

* Corresponding author. Fax: +31 71 526 6761

E-mail address: A.E.Gorbalenya@lumc.nl (A.E. Gorbalenya).

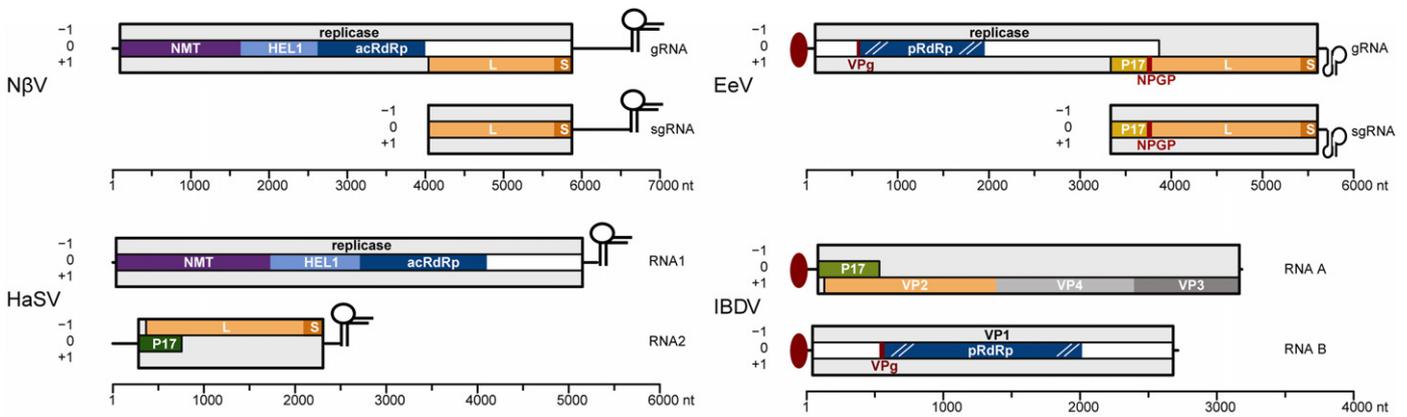


Fig. 1. Tetravirus and birnavirus genome comparisons. The genome organization, including genome segments, ORFs, selected domains and terminal structures, are depicted for three tetraviruses EeV, N β V, and HaSV and the IBDV birnavirus, using virus-specific scales. EeV, and N β V (*Betatetraviruses*) have monopartite genomes (g) that (are predicted to) yield sgRNAs (see text), while HaSV (*Omegetetraviruses*) have bipartite genomes (RNA1 and RNA2), as do birnaviruses (RNA A and RNA B). The 3'-tRNA-like structures of the N β V and HaSV genomes and the pseudoknots in EeV genomes/subgenome are indicated. Likewise, the (putative) VPg is shown at the 5'-end of RNAs in EeV and IBDV. The selected proteins and domains discussed in the text are labeled, and pattern-coded and colored to indicate homology. Note that P17s of three viruses are *not* homologous. The RF areas that are not known to encode proteins are colored in gray (light gray in the black-and-white version of figure) in all genomes. The figure is a manually refined version of a draft prepared using a genome drawing program written in R (CL and AEG, unpublished).

compact than those from betatetraviruses and display neither pits, nor grooves (Johnson and Reddy, 1998).

In all tetraviruses, the capsid precursor protein undergoes post-assembly autoproteolytic processing to yield two major capsid proteins. First shown to be an assembly-dependent maturation cleavage in N ω V (Agrawal and Johnson, 1992, 1995; Canady et al., 2000), this processing has also been observed in N β V (Gordon et al., 1999), HaSV (Hanzlik et al., 1995), *Dendrolimus punctatus* tetravirus (DpTV) (Yi et al., 2005), *Thoesa asigna* virus (TaV) (Pringle et al., 1999, 2001) and *Providence* virus (PrV) (Pringle et al., 2003). Additionally in TaV and PrV, the capsid precursor undergoes an apparent processing event immediately upstream of the L protein during RNA translation. It is autocatalytically mediated by the NPGP motif (Pringle et al., 2001; Luke et al., 2008), originally identified in 2A proteins of some picornaviruses (Donnelly et al., 2001). 2A was shown to promote an internal translation termination that is followed by elongation of translation of mRNA (Doronina et al., 2008). In PrV, tandem NPGP motifs may be operational (Luke et al., 2008).

The tetraviruses with single processing of the capsid precursor also share a conserved domain organization of the replicase gene that includes putative N7-methyltransferase (NMT), superfamily 1 helicase (Hel1) and RNA-dependent RNA polymerase (acRdRp) domains typical of the vast Alphavirus-like virus supergroup (Ahluquist et al., 1985; Gordon et al., 1995). In contrast, TaV replicase has no NMT and Hel1 domains but includes a structurally unique RdRp with the permuted active site (pRdRp). It was originally identified only in the EeV/TaV RdRp and the homologous domain of virus protein 1 (VP1) of the dsRNA *Birnaviridae* (Gorbalenya et al., 2002; Garriga et al., 2007; Pan et al., 2007). Importantly, pRdRps of these two virus families are likely to be monophyletic (Gorbalenya et al., 2002). Thus, TaV was speculated to prototype a unique lineage of ssRNA+ viruses with a mosaic relationship to birnaviruses (replicase) and prototypic tetraviruses (capsid) (Gorbalenya et al., 2002). Subsequent structural analysis of the capsid of the birnavirus infectious bronchitis disease virus (IBDV) (Coulibaly et al., 2005) showed that major capsid proteins of tetraviruses and birnaviruses are also related through a common jelly-roll β barrel fold (Munshi et al., 1998) that is wedge-shaped in both proteins. The sequence of the replicase gene of PrV, another tetravirus with complex processing of the capsid precursor (Pringle et al., 2003), suggests it may belong to another phylogenetic lineage (Pringle and Ball, personal communication).

In this report, we present the characterization of the capsid and complete sequence of the genomic RNA of *Euprosterina elaeasa* virus

(EeV), a promising candidate for biological control of its insect larval host *E. elaeasa*, Dyar (Lepidoptera: Limacodidae), a major defoliating pest of palm-oil trees in Central and South America (Genty, 1976; Genty et al., 1978). EeV is closely related to TaV and shares signature properties distinguishing these viruses from prototypic tetraviruses, including a predicted conserved pseudoknot at the genome 3'-end. Bioinformatics analysis of TaV/EeV replicases and VP1 of birnaviruses revealed an unprecedented conservation of a newly defined VPg signal upstream of the pRdRp domain. Phylogenetic analysis of the jelly-roll capsid domain in *Tetraviridae* rooted with the *Birnaviridae* was used to propose an evolutionary scenario that accounts for the monophyletic origin of the T = 4 capsid, the extraordinary diversity of replicases in tetraviruses, and the emergence of bipartite tetraviruses during co-infection of two distantly related monopartite tetraviruses.

Results

Morphology and protein composition of EeV particles

Examination of purified EeV particles using electron microscopy reveals non-enveloped isometric virions of 40 nm diameter (Fig. 2A),

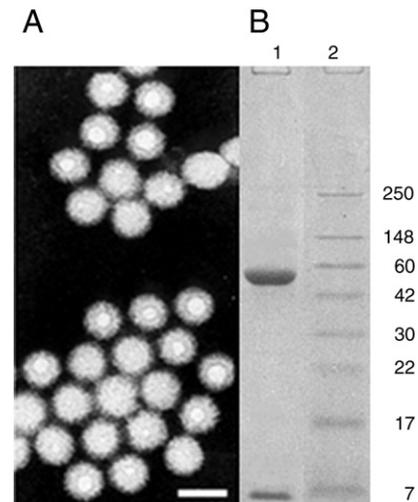


Fig. 2. Purified EeV particles: morphology and proteins. (A) Transmission electron micrograph (bar, 50 nm), and (B) SDS-PAGE analysis of purified EeV particles (lane 1). Markers (lane 2) are in kDa.

consistent with the morphology and size of tetraviruses (Hanzlik et al., 2005; Hanzlik and Gordon, 1997). SDS-PAGE analysis identified two capsid proteins, a major protein of approximately 58 kDa and a minor protein of approximately 7 kDa (Fig. 2B), within the range and relative intensities found for other tetraviruses: TaV (56 and 6 kDa) (Pringle et al., 1999), NβV (61 and 7.9 kDa) (Hanzlik and Gordon, 1997; Hanzlik and Gordon, 1998), NωV (62 and 7.8 kDa) (Agrawal and Johnson, 1992), PrV (60 and 7.4 kDa) (Pringle et al., 2003) and DpTV (62.5 and 6.8 kDa) (Yi et al., 2005).

EeV genome organization

More than 80% of the nucleotide sequence of the EeV genome was determined from four different overlapping cDNA clones. Five 5' RACE clones terminated with the identical 5'-AGGTCCACCTTCTTGTA sequence that is considered likely to include the very 5'-end. The 5'-terminal sequence was verified by direct sequencing of PCR products to exclude a possibility that it was a cloning artifact. It was concluded that this is the 5'-terminal sequence of the EeV genome.

cDNAs containing the 3'-end of the genome were synthesized by 3' RLM-RACE, then cloned and sequenced to establish that the 3'-terminal sequence is AAAUCCUUUCCACGCG-3'. Although no poly (A) tail was found, it was possible to prime cDNA synthesis with oligo (dT) due to an internal (A)-rich zone (starting at nt 3323).

The EeV genome contains a total of 5698 nucleotides (GenBank AF461742), composed of 24.3% A, 23.6% U, 23.8% C and 28.3% G. It has untranslated 5'- and 3'-terminal regions of 94 and 92 nts respectively. There are two major ORFs (Fig. 1). ORF1 (nts 95–3868) encodes a 1257 aa protein with a predicted molecular mass of 140.5 kDa. ORF2 is located at the 3'-end of the genome, overlapping ORF1 in a + 1 frame by 520 nucleotides. The AUG initiation codon for the EeV and TaV capsid precursors has not been determined experimentally. While the first AUG triplets at the start of ORF2 of these viruses are located two codons apart in a genome-wide alignment (data not shown), the second AUGs along with adjacent nucleotides are conserved and also in a better context (EeV: AGAatgA; TaV: AAAatgA) for translation initiation (Ranjan and Hasnain, 1995). Translation from this first conserved AUG codon would produce a protein of 752 aa with a calculated molecular mass of 82.5 kDa (nucleotides 3348–5606). No ORF larger than 291 nt was detected in the negative-sense RNA.

MFOLD analysis identified a 5'-terminal hairpin (nt 7–28) of -25.2 kcal/mol with a GUAU loop flanked by a 9 nt-long inverted repeat (Fig. 3A). Use of the PLMM_DPSS algorithm suggested a possible 5' pseudoknot that was not confirmed by further analysis using pknotsRG and Kinefold (data not shown).

At the 3'-termini of EeV/TaV genomic RNAs, no tRNA-like structure, as found in the prototypic tetraviruses, could be identified using programs such as RNAfold, consistent with lack of a 3'-terminal CCA.

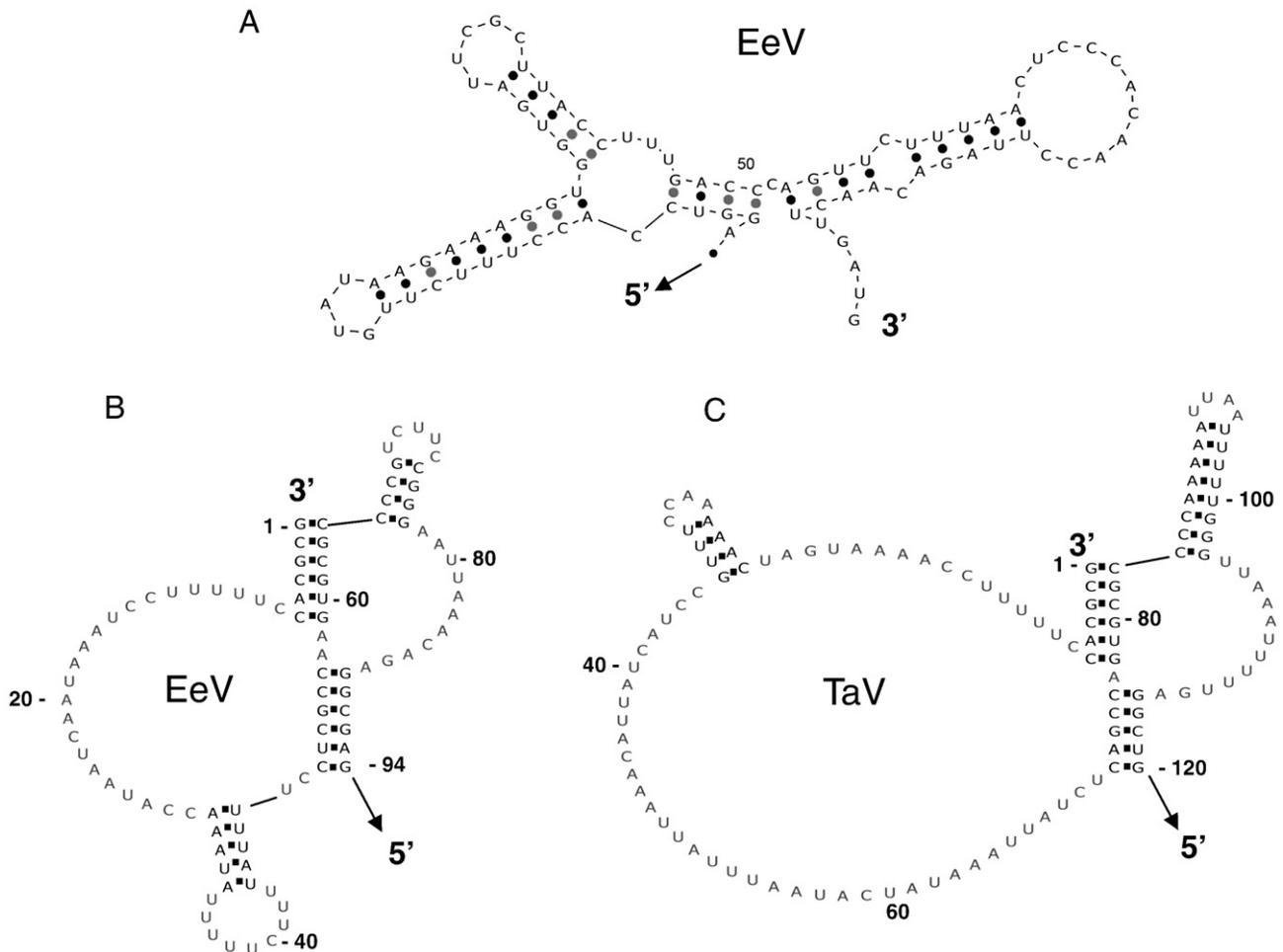


Fig. 3. Folding of the genomic RNA termini in EeV and TaV. (A) MFOLD prediction for folding of the EeV 5'-UTR. (B) RNA secondary structure models for the 3'-ends of the EeV and (C) TaV RNAs. Only the 3'-terminal 94 (EeV) or 120 (TaV) nucleotides are shown, numbered from the 3'-end. The structures were drawn using the PseudoViewer program of Byun and Han (2009) (URL: <http://wilab.inha.ac.kr/pseudoviewer/>). Base-pairs in the pseudoknot and other stems are indicated by dots. The short stem-loop formed by nts 65–77 (EeV) or nts 84–103 (TaV) is shown to the right of and above the pseudoknot, but is likely stacked with the pseudoknot.

The conservation of the TaV and EeV 3'-terminal sequences is mostly evident between (counting from the 5'-end) the first 19 nt of EeV and 49 of TaV, as well as between nt 51–94 (EeV) and nt 72–120 (TaV). Structurally related pseudoknots with comparable free energies of -22.50 and -24.90 kcal/mol respectively were tentatively identified in EeV and TaV (Figs. 3B and C) using pknotsRG and Kinfold. The pseudoknots contain very similar sequences that form two stems of comparable sizes. They are predicted to have the short conserved stem-loop located from 65/84 to 77/103 nt from the 3'-end stacked co-axially. The sequences containing the mainly unstructured A&U-rich loops differ in length (nt 7–50 for EeV; nt 7–71 for TaV) and may contain short, virus-specific, stem-loops. The predicted structures were also evident (unpublished observations) in analyses of larger 3'-terminal sequences than those depicted in Figs. 3B and C.

The EeV replicase: delineation of a nucleotidylation (VPg) signal upstream of the permuted RdRp

The 1257 aa EeV replicase ORF was used in our previous analysis of the identically sized TaV replicase (Gorbalenya et al., 2002), with which it has 68% aa identity also over the pRdRp domain that is conserved in the VP1 of the *Birnaviridae* (Fig. 1). The pRdRp domain commences at ~200 aa from the N-termini of the TaV/EeV replicases and VP1 of birnaviruses (Gorbalenya et al., 2002). We have now used profile-to-profile analysis to achieve a multiple sequence alignment that extends the remote conservation between TaV/EeV and birnaviruses to the very N-terminus (Fig. 4). Notably, there is a conserved position that includes a Ser residue whose hydroxyl group is guanylated *in vitro* in a fraction of VP1 molecules in infectious pancreatic necrosis virus (IPNV) (Xu et al., 2004). The Ser residue is replaced by a physico-chemically similar Thr residue in some birnaviruses and TaV (Fig. 4); this replacement introduced in the IPNV VP1 was shown to be compatible with the guanylation (Xu et al., 2004). The nucleotidylated Ser/Thr residue is part of the highly conserved Y/FXXGS/TXXGXXXR signature (dubbed VPg motif) (Fig. 4), whose conserved residues may contribute to the nucleotidylation reaction. In the IBDV VP1 structure, the VPg motif encompasses the interdomain junction and adjacent residues from the N-terminal domain and the most N-terminal α -helix 5 of the pRdRp (Garriga et al., 2007).

Upstream of the VPg motif, four α -helices and four β -strands were identified in the IBDV VP1 structure with the N-terminal 26 aa residues remaining unresolved (Garriga et al., 2007). The conservation in this area is only moderate among birnaviruses (Fig. 4). Consistently, only two Pro and one Asn have been found to be invariant among TaV/EeV

and birnaviruses; variation of dozens of other residues is limited within groups of physico-chemically similar residues (Fig. 4). These observations indicate that the N-terminal domain may adopt a similar structure and have a similar function in TaV/EeV and birnaviruses, although secondary structure prediction for TaV/EeV supported only the presence of the conserved α -helices in this domain (data not shown).

The sequences C-terminal to the pRdRp domain comprise ~600 and ~300 aa residues in the TaV/EeV replicase and birnavirus VP1, respectively. In the IBDV VP1, this domain adopts a unique α -helix fold that was implicated in the regulation of the RdRp activity (Garriga et al., 2007). No appreciable across-family similarity was found in this region.

Two-site capsid precursor processing pathway is conserved in EeV and TaV

Initiation of the capsid precursor synthesis at the first conserved AUG codon of EeV ORF2 would yield a protein of 82.5 kDa. Based on the strong similarity with TaV (Pringle et al., 1999, 2001; Donnelly et al., 2001), three proteins are predicted to be derived from this precursor, due to translation interruption mediated by the NPGP motif (between Gly-146 and Pro-147 of the motif) and assembly-dependent, autocatalytic cleavage between Asn-701 and Gly-702 (Taylor and Johnson, 2005). These proteins are (from the N- to C-terminus) P17 (146 aa), L (544 aa) and S (62 aa), the latter two forming the mature capsid. SDS-PAGE analysis of EeV particles showed two major structural proteins of 58 kDa (L) and 7 kDa (S) (Fig. 2B). N-terminal sequencing of these proteins yielded the sequences PPSVARGLQE and GWGLM, matching the deduced sequences at 147–156 and 698–702 in the EeV capsid precursor, respectively, and confirming the predicted sites. Blast searches revealed that the predicted and similarly positioned EeV and TaV P17s (Fig. 1) are orthologous (51.0% identity, 63.2% similarity) but otherwise have no similarity with other proteins outside a short region linked to the NPGP motif (Pringle et al., 2001). These results strongly support the conservation of the two-step processing of the capsid precursor in EeV and TaV that distinguishes these viruses from most tetraviruses.

The rooted phylogeny of tetraviruses

According to psi-BLAST-mediated analysis, capsid proteins of tetraviruses form a separate cluster, with the EeV and TaV capsid precursors being most closely related (18–87% identity range for the

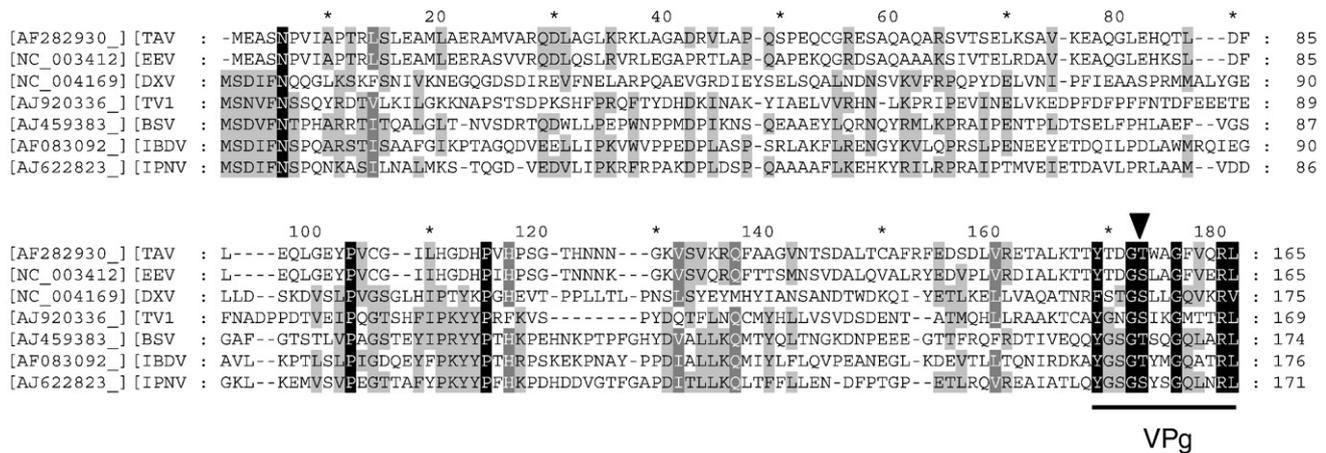


Fig. 4. Conservation in the N-terminal part of replicases of tetraviruses and birnaviruses. Muscle-based alignment of the N-terminal region of replicases TaV and EeV (top two rows) and VP1 of five birnaviruses, representing major lineages, is depicted. Sequence names are extended to include respective accession numbers in GenBank/RefSeq databases. Invariant and conserved residues are shadowed accordingly; the VPg motif is underscored and the predicted nucleotidylated Ser/Thr residue is indicated with a solid triangle. This motif is part of the 28 aa hit (compass, $E = 10^{-2}$) between TaV/EeV and birnaviruses.

Table 1

Pairwise maximum likelihood distances (bottom number) and percentage identities (top number) for capsid proteins of the *Tetraviridae* (bottom left triangle) and replicases of prototypic tetraviruses (upper right triangle)^a.

Replicase capsid	PrV	HaSV	N ω V	DpTV	N β V	EeV	TaV
PrV		n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
HaSV	35		n.a.	69	33	n.a.	n.a.
	1.66			0.45	1.96		
N ω V	34	66		n.a.	n.a.	n.a.	n.a.
	1.74	0.50					
DpTV	33	67	87		33	n.a.	n.a.
	1.75	0.47	0.15		1.95		
N β V	18	20	20	21		n.a.	n.a.
	3.63	3.28	3.27	3.28			
EeV	19	22	21	20	32		82
	3.46	3.01	3.17	3.22	1.86		0.22
TaV	18	23	21	22	33	84	
	3.58	2.98	3.14	3.17	1.87	0.19	

^a ML distances estimate the average number of substitutions per amino acid position. Percentage distances were calculated for the jelly-roll domain (capsid proteins), concatenated NMT, HEL1 and acRdRp domains (replicase proteins with the domain layout conserved in Alpha-like viruses) or pRdRp domain (TaV and EeV). Bold numbers are discussed in text. n.a., not available.

family) (Table 1). To study tetravirus evolution, we generated a DaliLite-mediated structurally based sequence alignment between N ω V L and IBDV VP2 proteins (Helgstrand et al., 2004; Coulibaly et al., 2005), which were found to be most structurally similar outside tetraviruses (17SD DaliLight score). The linear conservation of structural elements between the base (B) and S domains of IBDV and the L protein of N ω V is shown diagrammatically in Fig. 5A. Both the IBDV S and N ω V L domains include structural elements of the jelly-roll fold that are uniquely interspersed with domains that are conserved in either birnaviruses (the IBDV P domain) or tetraviruses (N ω V Ig-like domain). Using the profile-to-sequence mode of Clustal W and Muscle programs, this alignment has been extended to orthologs from other species of these two families, including TaV and EeV.

A multiple capsid alignment of seven tetraviruses was used to infer trees rooted with five birnaviruses. Using either BEAST-mediated Bayesian inference or PAUP-mediated parsimonious analysis, tetravirus capsid trees with identical topologies were reconstructed (Figs. 5B and C). The internal branching in the Bayesian tree was fully supported (Fig. 5B). In contrast, topologies of the birnavirus trunk differed in the two analyses. Accordingly, the internal branching of the Bayesian tree was not uniformly supported. This uncertainty is beyond the scope of this paper and does not affect its conclusions.

The tetravirus capsid tree includes the two branches that are formed by the monopartite genome betatetravirus N β V and its sister group (EeV/TaV), respectively, and the other two branches leading to the monopartite genome PrV and three bipartite omegatetraviruses, HaSV, N ω V and DpTV (Fig. 5B and C), respectively. Monopartite viruses employ different, branch-specific replicases, one of which with the NMT-Hel1-acRdRp domain layout is also and uniformly used by bipartite tetraviruses (Figs. 1 and 5B and C) (prototypic tetraviruses). A Bayesian tree of the replicase of prototypic tetraviruses was inferred with high confidence and rooted with the most closely related hepeviruses (Gordon et al., 1995). It has a topology matching that of the capsid tree of this subset of tetraviruses (Fig. 5B) with a notable specific: mono- and bipartite prototypic tetraviruses are monophyletic in the replicase tree but may not be so in the capsid tree, in which they cluster with TaV/EeV and PrV lineages, respectively. These peculiarities and relatively large distances separating mono- and bipartite prototypic tetraviruses compared to those observed within the bipartite *Omegatetraviruses* (0.15–0.50 vs. 1.95–3.28) (Table 1) indicate bipartite viruses to have originated late in the evolution. Remarkably, while ML distances separating diverse bipartite tetraviruses in the most conserved domains of capsid and replicase are

comparable (0.47 and 0.50 vs. 0.45), those separating mono- to bipartite tetraviruses in these domains are not, with the capsids compared to replicases having accepted ~1.67 times more substitutions (3.28 and 3.27 vs. 1.96) (Table 1, bold numbers). This difference could be due to either accelerated evolution of the capsid or recombination involving non-prototypic tetraviruses upon the transition from mono- to bipartite tetraviruses.

Discussion

This study describes the genome sequence, virion composition and morphology of EeV, a second ssRNA+ virus that employs a pRdRp. EeV is shown to have a number of distinct characteristics in common with the closely related TaV. These features, including the permuted replicase as previously described (Gorbalenya et al., 2002), related capsid expression strategies, a predicted 3'-pseudoknot structure on the viral RNA and a nucleotidylase signal in the replicase, all place TaV/EeV apart from other vastly different viruses within the family *Tetraviridae*, that is defined by T=4 capsid symmetry. Using phylogenetic analysis of the *Tetraviridae* we show how this diversity might have evolved.

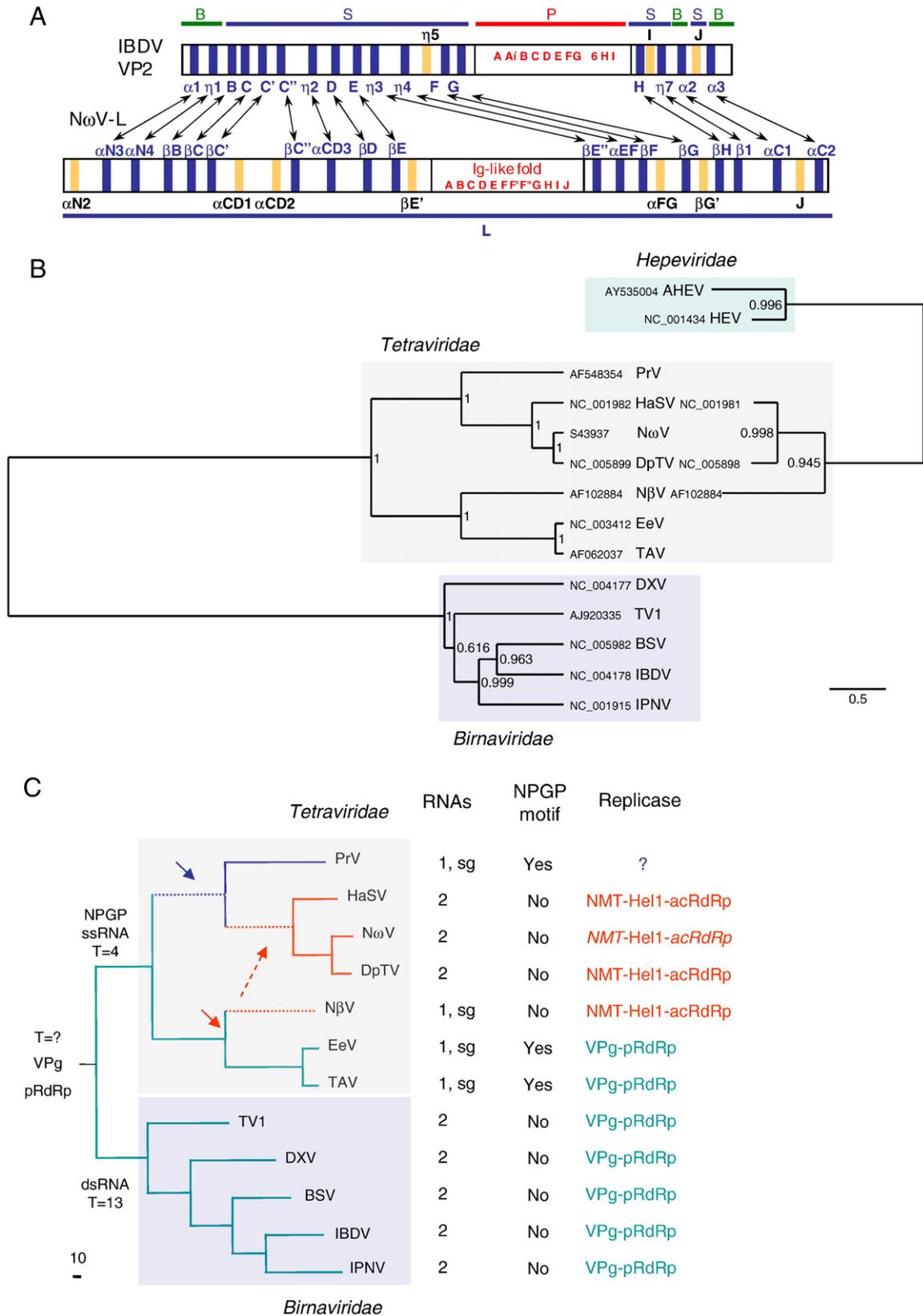
Comparison of EeV and TaV RNA terminal sequences, based on analysis of multiple clones, shows them to be very similar. EeV possesses an extra eight 5'-nucleotides, suggesting that the sequence of TaV (Gorbalenya et al., 2002) was not completely determined. Bioinformatics analysis further predicts that the 5'-end of TaV/EeV is blocked with a protein (see below), a prediction that remains to be verified experimentally. The TaV/EeV 3'-terminus is predicted to form a conserved pseudoknot with the 3'-terminal residue located at its base. Pseudoknots are found in the 3'-non-coding regions of many RNA viruses, and may be involved in regulation of genome replication (Brierley et al., 2007). The most relevant to TaV/EeV may be the pseudoknot of tomato bushy stunt virus that has the 3'-terminus buried within stacked helices (Na et al., 2006). This "closed" conformation is thought to prevent genome replication, favoring translation. An alternative, "open", conformation disrupts the pseudoknot to enable replication to occur. This model might be relevant in regulation of EeV/TaV genome expression and replication. The predicted 3'-end pseudoknot may also or alternatively protect EeV/TaV RNAs from exonucleases.

The differences in the predicted terminal structures for EeV/TaV genomic RNAs from those of prototypic tetraviruses, that are predicted to have a 5'-cap and a 3'-tRNA-like structure (Hanzlik and Gordon, 1997), correlate with the differing domain compositions of replicases in these viruses. Replicases of prototypic tetraviruses include the NMT-Hel1-acRdRp array of conserved domains; in contrast, TaV/EeV replicases have no Hel and MT domains and encode a non-canonical pRdRp in the N-terminal half (Fig. 1). The lack of a Hel domain in TaV/EeV, and its presence in prototypic tetraviruses, are consistent with the observation that ssRNA+ viruses with a genome size above 6–7 kb have a Hel domain, whereas those under 6–7 kb do not (Gorbalenya and Koonin, 1989). The TaV/EeV and birnavirus replicases share an additional conserved N-terminal domain not found in prototypic tetraviruses. It includes a newly described VPg motif used in birnaviruses to covalently link 5'-ends of positive-strand RNA of two genomic segments to a fraction of VP1 molecules (replicase), thought to serve as protein primers for replication (Dobos, 1995; Magyar et al., 1998; Xu et al., 2004). Likewise, this signal could be used by TaV/EeV to prime genomic and sgRNA synthesis; since genomic RNA of EeV starts with an adenine, this signal is predicted to be adenylated. The priming of RNA synthesis with a protein in TaV/EeV would be consistent with the distant phylogenetic affinity of pRdRp and canonical RdRps of other RNA viruses that use VPg (Gorbalenya et al., 2002; Hanzlik et al., 2005; Garriga et al., 2007).

The across-family conservation of the VPg signal, evident in TaV/EeV and birnaviruses and unprecedented for other VPg types

in many RNA virus families (Le Gall et al., 2008; Gorbalenya, unpublished), further supports the evolutionary relationship between replicases of TaV/EeV and birnaviruses spanning across two different RNA virus classes (Gorbalenya et al., 2002). The special affinity between two families also extends to the capsid and also

includes tetraviruses that employ different replicases (Coulbaly et al., 2005). A distinct structural variant of the jelly-roll domain was found to be used for building the T=4 capsid of tetraviruses and the external shell layer of the T=13 capsid of birnaviruses. We used this observation to root the tetravirus capsid tree in our study



(Figs. 5B and C). As argued previously (Coulibaly et al., 2005), the ancestral virus for tetra- and birnaviruses might have used the $T=3$ virions evident in nodaviruses (Schneemann et al., 1998) or $T=4$ virions. Alternatively if the ancestor was very small it might have employed $T=1$ virions; strikingly, this latter arrangement is evident in subviral particles formed by VP2 expressed alone (Caston et al., 2001; Coulibaly et al., 2005). This ancestor was possibly a ssRNA+ virus since dsRNA birnaviruses most resemble viruses with ssRNA+ genomes (Gorbalenya et al., 2002; Coulibaly et al., 2005; Ahlquist, 2005). Most recent common ancestors (MRCA) of birnaviruses and tetraviruses are separated by a very large distance exceeding those observed within each family (Fig. 5B); accordingly, multiple family-specific insertions are evident in the jelly-roll domains of viruses of these two families (Fig. 5A). Recent publication of the genome sequence and characterization of *Drosophila A* virus (DAV) (Ambrose et al., 2009) shows that viruses of newly emerging families may employ pRdRp and jelly-roll-based capsid. The DAV replicase also includes a VPg motif upstream of the pRdRp (AEG, unpublished observation). To accommodate DAV, which is unlikely to be the last virus of this growing diverse group, further revision of the above evolutionary scenario might be necessary.

The replicase of the virus with this ancestral capsid probably employed a pRdRp, since viruses with a VPg-pRdRp are found in both birna- and tetraviruses (Fig. 5C). Likewise, the ancestral tetraviruses might, like PrV, N β V and TaV/EeV and in contrast to phylogenetically compact omegatetraviruses, have had a monopartite genome. The phylogenetic analysis further suggests that, the replicase with pRdRp was replaced independently in the two lineages leading to PrV and N β V by, respectively, a replicase of an unknown type and a replicase with the NMT-Hel1-acRdRp domain layout. This reconstruction highlights the conundrum presented by the omegatetraviruses: while these viruses derive their replicase from an N β V-like ancestor, their capsid proteins are more closely related to that of PrV. These viruses are also distinctive among tetraviruses in expressing the capsid gene from a separate genomic RNA rather than from a sgRNA.

Two evolutionary scenarios for the origin of the bipartite omegatetraviruses, accounting for major specific features of tetraviruses, can be drawn. Both exploit a unique property of diverse monopartite tetraviruses, all of which have been shown or are assumed to encapsidate a sgRNA along with the genomic RNA. (Evidence for sgRNA in the EeV virions has not been obtained due to limited quantities of the specimen available for the analysis.) They require co-infection of a cell by two different parental monopartite tetraviruses of the N β V and PrV lineages but differ regarding important aspects of their interaction. According to the first scenario, the co-infection could have given rise to a hybrid virion progeny including the genomic RNA encoding an N β V-like replicase and a subgenomic RNA encoding a PrV-like capsid. Possible mutations severely compromising

the replicase-donor's capsid gene expression and enabling replication of a mutant PrV capsid sgRNA by the N β V-like replicase would have promoted subsequent proliferation of the progeny to establish a new bipartite genome lineage. Numerous concerted mutations are not uncommon in RNA viruses; they must have been involved in the RdRp diversification to generate pRdRp and canonical RdRp in a progenitor of this lineage (Gorbalenya et al., 2002) and in the independent emergence of pRdRp in a species of the Tymoviridae (Sabanadzovic et al., 2009). The above scenario exploits the natural evolutionary potential of a sgRNA toward becoming a genomic segment in a process resembling reassortment in segmented viruses that could be driven by genetic drift or selection. An alternative, and more complex, scenario postulates recombination between the parental viruses to generate an intermediate monopartite virus with N β V-like replicase and PrV-like capsid, followed by genome segmentation to separate these genes. There is no apparent rationale, however, why the process of genome segmentation required by this scenario (and which appears to have happened repeatedly among the ssRNA+ viruses (Strauss and Strauss, 1988; Le Gall et al., 2008), should have occurred in this putative lineage, when it has not been observed in any of the ancestral monopartite tetraviruses lineages, i.e. N β V, PrV and TaV/EeV lineages. We favor the first scenario.

Despite belonging to two different tetraviruses lineages, TaV/EeV and PrV share another conserved protein module, characterized by the ~17 aa NPGP motif (Pringle et al., 2003; Luke et al., 2008). Its small size makes reliable phylogenetic reconstruction challenging, although the clustering of TaV/EeV and PrV in a tree comparing the NPGP-containing proteins of ssRNA+ viruses has been reported (Luke et al., 2008). An alternative approach to infer the evolutionary history of this module is through analysis of associated capsid proteins. Assuming that the observed association is stable, the ancestral NPGP module might have already been present in the MRCA of the tetraviruses capsids. It was subsequently and independently lost in two lineages, leading to prototypic omega- and betatetraviruses, respectively (Fig. 5C). Alternatively, if the NPGP module is promiscuous, it could have been introduced from either the TaV/EeV or PrV lineage to the other lineage by recombination.

The family *Tetraviridae* clearly includes viruses with vastly incongruent evolutionary histories of capsid and RdRp proteins. This conflict, challenging for virus taxonomy, is not evident in most families of ssRNA+ viruses, the foundations of which are phylogenetically sound. To resolve this conflict, TaV/EeV and, probably PrV, must be placed in two new families separate from the prototypic tetraviruses. Since all these viruses possess a $T=4$ capsid, their family names could include "*Tetraviridae*" with a prefix referring to the unique replicase. Accordingly, we propose a family prototyped by TaV/EeV to be named *Permutotetraviridae* and the current *Tetraviridae*, encompassing prototypic viruses, be renamed *Alphatetraviridae* or split further to recognize prototypic *Beta-* and *Omegatetraviruses* as

Fig. 5. Capsid-based alignment and phylogenetic tree of tetraviruses and birnaviruses. (A) Schematically depicted is a DaliLite-mediated structural alignment of the NoV and IBDV capsid proteins. The β -sheet and α -helix structural elements that are conserved in the S and L proteins of IBDV and NoV, respectively, are shown as blue rectangles that are connected by double-headed arrows. Structural elements that have no counterpart in the other protein are labeled above IBDV VP2 or below NoV L proteins, respectively, or indicated in red (B and P domains of IBDV and Ig-like domain of NoV). (B) Phylogeny of tetraviruses. The capsid tree for seven tetraviruses is based on an amino acid alignment of the Jelly Roll domain (L protein) (661 positions) and was rooted using S protein of five birnaviruses as an outgroup (left side). The replicase tree for prototypic tetraviruses and DpTV is based on an amino acid alignment of the HEL1 and RdRp domains (832 positions) and was rooted using avian and human Hepatitis E viruses as an outgroup. NoV was not included in this and another phylogenetic analysis described in (C) since only the HEL1 domain has been sequenced for the replicase of this virus (Hanzlik et al., 2005); the available sequence is compatible with the tree (unpublished data). Numbers at branch points provide posterior probability support values. The trees were calculated under strict and relaxed lognormal molecular clock models, respectively. The scale of both phylogenies is indicated by the bar of 0.5 amino acid substitutions per site on average. GenBank/RefSeq accession numbers are indicated next to the virus names. Three virus families are highlighted with different color background. (C) Evolution of tetraviruses and birnaviruses in the context of the most parsimonious tree of capsid proteins. The most parsimonious tree by the exhaustive search of the entire tree-space (654,729,075 trees) was calculated for the alignment (see A and B) that includes 431 parsimony-informative characters. The tree has consistency index 0.9212 and rescaled consistency index 0.7736. Bar indicates the number of steps; the unrooted tree has 2120 steps. The tree was rooted on the branch connecting two virus families. Major characteristics of the families and possible characteristics of the common progenitor of two families are indicated at respective basal nodes. For contemporary viruses, the number of genomic and subgenomic RNAs used, replicase domain layout and the presence of the NPGP motif are indicated. The NoV replicase is depicted in italic to reflect the fact that it has been sequenced only partially (see B) (Hanzlik et al., 2005). Replicases were colored to visualize their identity and this coloring scheme was also applied to branches in the tree where respective viruses are either observed or expected. Dotted branches indicate ranges where a switch of replicase could have occurred. Also colors of arrows indicating replacement of the replicase in the virus lineage was matched to the color of incoming replicase. Solid arrow, replicase from an outside source; broken arrow, replicase from a paraphyletic tetraviruses lineage.

two families. The recognition of multiple virus families with T=4 architecture and monophyletic history of the major proteins would be somewhat reminiscent of relationships observed between numerous virus families with the pseudo T=3 architecture now united in the order Picornvirales (Le Gall et al., 2008). A similar rationale could also be adopted for consistent naming of other newly emerging RNA virus families whose members employ pRdRp and jelly-roll-based virions with non-T=4 symmetry.

Materials and methods

Source of virus-infected larvae and virus purification

Cadavers of *E. elaeasa* larvae were collected in 1996 from the Peruvian palm-oil plantation (Palmas del Espino, near Uchiza) after natural epizootics and stored at -20 °C. EeV particles were extracted from the frozen cadavers and purified on sucrose gradients as described by Hanzlik et al. (1993).

Transmission electron microscopy

Gradient purified virus was deposited on 200-mesh grids and stained with 2 % phosphotungstic acid according to Brenner and Horne (1959). Observations were made with a Zeiss EM-10-CR transmission electron microscope.

N-terminal sequencing of EeV capsid proteins

Virus samples were denatured then subjected to electrophoresis on 12.5% SDS-polyacrylamide gels (SDS-PAGE) (Laemmli, 1970). After staining with Coomassie blue, the capsid protein bands were excised and subjected to 10 cycles of N-terminal sequencing by Edman degradation at the Australian National University Biomolecular Resource Facility.

cDNA synthesis and cloning

EeV RNA was extracted from purified capsids as described by Hanzlik et al. (1993). cDNA was synthesized in the presence of random hexamers using Superscript II reverse-transcriptase (Life Technologies). The cDNAs were blunt-ended with T4 DNA polymerase (New England BioLabs), size selected for lengths >1 kbp on a spin column (Clontech) and cloned into the *EcoRV* site of pBluescript (SK⁺) (pBS, Stratagene) using standard procedures.

Determination of terminal sequences

A 5'-RACE kit (Life Technologies) was used to amplify the 5' ends according to the manufacturer's specifications with minor modifications. cDNAs synthesized with Superscript II and the EeV-specific primer 450R (complementary to nucleotides 1132 to 1154: 5'-TCTGTATCATCACCATAACAGGC-3') were dC-tailed using terminal deoxynucleotidyl transferase (TdT). After annealing to Abridged Anchor Primer (AAP: 5'-GGCCACGCGTCTGACTAGTACGGGIIIGGGIIGGG IIG-3'; Life Technologies), dC-tailed cDNAs were subjected to 20 cycles of PCR (extension time: 3 min) using Pfu DNA polymerase (New England BioLabs) and primer 181R (complementary to nucleotides 397 to 420: 5'-GTGCCGAGGGGTGGATAGGATGG-3'). PCR products recovered from agarose gels were directly sequenced, or phosphorylated using T4 polynucleotide kinase (New England Biolabs), cloned into pBS, and sequenced. Mapping of the viral RNA 3'-terminus was achieved using T4-RNA ligase mediated amplification of cDNA ends or RLM (RNA ligase mediated) -RACE (Liu and Gorovskiy, 1993). Phosphorylated EeV3R primer (5'-GTACTAGTCCGCGTGCC (TAG)₄-3') was ligated to EeV RNA with 100 units of T4 RNA ligase (New England Biolabs). cDNA synthesis was primed with Abridged

Universal Amplification Primer (AUAP: GGCCACGCGTCTGACTAGTAC; Life Technologies) whose sequence is complementary to the 5'-end of that of EeV3R primer. The EeV-specific primer 4725F (NT 5411 to 5437: 5'-TATAATGGGTGGGGCTTGTATGTTCTC-3') was used conjointly with UAP ([CUA]₄GGCCACGCGTCTGACTAGTAC) for the PCR amplification of 3' terminus sequences. PCR products were resolved on agarose gels, cloned into pBS and sequenced.

Sequence analyses

cDNA clones were sequenced with an Applied Biosystems model 373A sequencer using the *Taq* dideoxy terminator sequencing. Universal forward and reverse primers were used in conjunction with synthetic primers based on sequenced portions of the virus. Genomic sequences were assembled using the GCG 8.1 software package (Genetics Computer Group, Madison, WI, USA (Devereux et al., 1984).

Bioinformatics analyses

RNA secondary structure was analyzed using MFOLD software (Zuker, 2003), Pseudoknot Local Motif Model and Dynamic Partner Sequence Stacking (PLMM_DPSS) algorithm of Huang and Ali (2007), RNAfold from <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>, pknotsRG (Reeder and Giegerich, 2004; Reeder et al., 2007); server at: <http://bibiserv.techfak.uni-bielefeld.de/pknotsrg/submission.html> and Kinefold (Xayaphoummine et al., 2005); server at <http://kinefold.curie.fr/cgi-bin/form.pl>. Amino acid (aa) sequence alignments were generated using either Megalign software (Lasergene; DNASTar Inc.) or Clustal (versions 1.81 and 1.82) (Thompson et al., 1994) and Muscle (Edgar, 2004) programs assisted by Blosum position-specific matrices (Henikoff and Henikoff, 1994) in the Vivalis software platform (Gorbalenya, unpublished), and they were processed for presentation using GeneDoc (Nicholas et al., 1997). Profiles were built from multiple sequence alignments using Compass (Sadreyev and Grishin, 2003). GenBank (Benson et al., 2008) was searched in default mode, unless otherwise stated, using FASTA (GCG) and Blast programs (Altschul et al., 1997). Pairwise comparison of protein structures was performed using DaliLite software (Holm and Park, 2000). Protein secondary structure predictions were made using PsiPred (Jones, 1999).

Phylogenetic analyses

Multiple amino acid sequence alignments of the capsid and replicase proteins were used for phylogenetic analysis employing the neighbor-joining (NJ) method in PHYLIP (Phylogeny Inference Package) version 3.5c (Felsenstein, 1989), parsimonious method with exhaustive search of the entire tree-space as implemented in the UNIX version of the PAUP* 4.0.b10 program (Swofford, 2000), and a Bayesian posterior probability approach utilizing the BEAST software (Drummond and Rambaut, 2007). For the latter analysis, MCMC chains (two per dataset) were run for 2 million steps (10% burn-in, sampled every 50 generations). For each Bayesian analysis three molecular clock models (strict, relaxed with lognormal distribution, relaxed with exponential distribution) were tested (Drummond et al., 2006). The more complex model, e.g. relaxed molecular clock, was favored over the simpler model, e.g. strict molecular clock, if the Bayes factor (ratio of tree likelihoods) was bigger than five (Goodman, 1999). Convergence of runs was verified and Bayes factors were estimated using Tracer (Rambaut and Drummond, 2007). Pairwise evolutionary distances (estimated number of amino acid substitutions per site on average) were calculated by applying a maximum likelihood (ML) approach using the Tree-Puzzle program (Schmidt et al., 2002). The WAG amino acid substitution model (Whelan and Goldman, 2001) was used and rate heterogeneity among sites

(gamma distribution with 8 categories) was allowed in the Bayesian and ML analyses.

Sequences used in bioinformatics analyses

GenBank/RefSeq accession numbers of virus sequences analyzed were: HaSV: [NC_001982, NC_001981], N β V: [AF102884], N ω V: [S43937], PrV: [AF548354], TaV: [AF062037, AF282930], DpTV: [NC_005899, NC_005898], *Infectious bursal disease virus* (IBDV): [NC_004178, AF083092], [NC_001915, AJ622823] for *Infectious pancreatic necrosis virus* (IPNV), *Blotched snakehead virus* (BSV): [NC_005982, AJ459383], *Tellina virus 1* (TV1): [AJ920335, AJ920336], *Drosophila X virus* (DXV): [NC_004177, NC_004169], Human and avian Hepatitis E viruses (HEV, AHEV): [NC_001434, AY535004]. Sequence names used in the alignment and trees were prepared using SNAD (Sidorov et al., 2009) at <http://veb.lumc.nl/SNAD/>.

Acknowledgments

We thank Michelle Williams and Susan Dorrian for technical assistance, Fiona Pringle for her contribution to this research and her comments on the manuscript, Andrew Ball and Fiona Pringle for sharing their unpublished data on PrV, and Alexander Kravchenko, Dmitry Samborsky and Igor Sidorov for administrating Viralis software. This work was partially supported by the New Zealand Marsden Fund, Contract UO0212 (WV) and the Netherlands Bioinformatics Center BioRange SP 3.2.2 grant (AEG).

References

- Agrawal, D.K., Johnson, J.E., 1992. Sequence and analysis of the capsid protein of Nudaurelia-capensis omega-virus, an insect virus with T = 4 icosahedral symmetry. *Virology* 190, 806–814.
- Agrawal, D.K., Johnson, J.E., 1995. Assembly of the T = 4 Nudaurelia capensis omega virus capsid protein, posttranslational cleavage, and specific encapsidation of its messenger-RNA in a baculovirus expression system. *Virology* 207, 89–97.
- Ahlquist, P., 2005. Virus evolution: fitting lifestyles to a T. *Curr. Biol.* 15, R465–R467.
- Ahlquist, P., Strauss, E.G., Rice, C.M., Strauss, J.H., Haseloff, J., Zimmern, D., 1985. Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. *J. Virol.* 53, 536–542.
- 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, 3389–3402.
- Ambrose, R.L., Lander, G.C., Maaty, W.S., Bothner, B., Johnson, J.E., Johnson, K.N., 2009. *Drosophila A virus* is an unusual RNA virus with a T = 3 icosahedral core and permuted RNA-dependent RNA polymerase. *J. Gen. Virol.* 90, 2191–2200.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2008. GenBank. *Nucleic Acids Res.* 36, D25–D30.
- Brenner, S., Horne, R.W., 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta* 34, 103–110.
- Brierley, I., Pennell, S., Gilbert, R.J.C., 2007. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat. Rev. Microbiol.* 5, 598–610.
- Byun, Y., Han, K., 2009. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics* 25, 1435–1437.
- Canady, M.A., Tihova, M., Hanzlik, T.N., Johnson, J.E., Yeager, M., 2000. Large conformational changes in the maturation of a simple RNA virus, Nudaurelia capensis omega virus (N omega V). *J. Mol. Biol.* 299, 573–584.
- Caston, J.R., Martinez-Torrecuadrada, J.L., Maraver, A., Lombardo, E., Rodriguez, J.F., Casal, J.I., Carrascosa, J.L., 2001. C terminus of infectious bursal disease virus major capsid protein VP2 is involved in definition of the T number for capsid assembly. *J. Virol.* 75, 10815–10828.
- Christian, P., Dorrian, S.J., Gordon, K.H.J., Hanzlik, T.N., 2001. Pathology and properties of the tetravirus *Helicoverpa armigera* stunt virus. *BioControl* 20, 65–75.
- Coulibaly, F., Chevalier, C., Gutsche, I., Pous, J., Navaza, J., Bressanelli, S., Delmas, B., Rey, F.A., 2005. The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell* 120, 761–772.
- Devereux, J., Haeblerli, P., Smithies, O., 1984. A comprehensive set of sequence programs for the VAX. *Nucleic Acids Res.* 12, 387–395.
- Dobos, P., 1995. Protein-primed RNA synthesis *in vitro* by the virion-associated RNA polymerase of infectious pancreatic necrosis virus. *Virology* 208, 19–25.
- Donnelly, M.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., Ryan, M.D., 2001. The ‘cleavage’ activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring ‘2A-like’ sequences. *J. Gen. Virol.* 82, 1027–1041.
- Doronina, V.A., Wu, C., de Felipe, P., Sachs, M.S., Ryan, M.D., Brown, J.D., 2008. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol. Cell. Biol.* 28, 4227–4239.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4, 699–710.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Entwistle, P.F., 1987. Virus diseases of Limacodidae. In: Cock, M.J.W., Godfray, H.C.J., Holloway, J.D. (Eds.), “Slug and Nettle Caterpillars: The Biology, Taxonomy, and Control of the Limacodidae of Economic Importance on Palms in Southeast Asia”. CAB International, Wallington, pp. 213–221.
- Felsenstein, 1989. PHYLIP—phylogeny interference package (version 3.2). *Cladistics* 5, 164–166.
- Finch, J.T., Crowther, R.A., Hendry, D.A., Struthers, J.K., 1974. Structure of Nudaurelia-capensis beta virus—first example of a capsid with icosahedral surface symmetry T = 4. *J. Gen. Virol.* 24, 191–200.
- Garriga, D., Navarro, A., Querol-Audi, J., Abaitua, F., Rodriguez, J.F., Verdaguier, N., 2007. Activation mechanism of a noncanonical RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20540–20545.
- Genty, P., 1976. Etude morphologique et biologique d’un lépidoptère défoliateur du palmier à huile en Amérique latine *Darna metaleuca*, Walker. *Oléagineux* 31, 99–107.
- Genty, P., Desmier de Chenon, R., Morin, J.P., Korytkowski, C.A., 1978. Oil palm pests in Latin America. *Oléagineux* 31, 325–419.
- Goodman, S.N., 1999. Toward evidence-based medical statistics: 2. The Bayes factor. *Ann. Intern. Med.* 130, 1005–1013.
- Gorbalenya, A.E., Koonin, E.V., 1989. Viral-proteins containing the purine Ntp-binding sequence pattern. *Nucleic Acids Res.* 17, 8413–8440.
- Gorbalenya, A.E., Pringle, F.M., Zeddiam, J.L., Luke, B.T., Cameron, C.E., Kalmakoff, J., Hanzlik, T.N., Gordon, K.H., Ward, V.K., 2002. The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J. Mol. Biol.* 324, 47–62.
- Gordon, K.H.J., Hanzlik, T.N., 1998. Tetraviruses. In: Miller, L.K., Ball, L.A. (Eds.), “The insect viruses”. Plenum Press, London, pp. 269–299.
- Gordon, K.H., Waterhouse, P.M., 2006. Small RNA viruses of insects: expression in plants and RNA silencing. *Adv. Virus Res.* 68, 459–502.
- Gordon, K.H.J., Johnson, K.N., Hanzlik, T.N., 1995. The larger genomic RNA of *Helicoverpa armigera* stunt tetravirus encodes the viral RNA polymerase and has a novel 3′-terminal tRNA-like structure. *Virology* 208, 84–98.
- Gordon, K.H.J., Williams, M.R., Hendry, D.A., Hanzlik, T.N., 1999. Sequence of the genomic RNA of Nudaurelia beta virus (*Tetraviridae*) defines a novel virus genome organization. *Virology* 258, 42–53.
- Hanzlik, T.N., Gordon, K.H.J., 1997. The Tetraviridae. *Adv. Virus Res.* 48, 101–168.
- Hanzlik, T.N., Gordon, K.H.J., 1998. Tetraviruses (*Tetraviridae*). In: Wrigglesworth, J. (Ed.), “The Encyclopedia of Virology”. Academic Press, New York, pp. 1764–1772.
- Hanzlik, T.N., Dorrian, S.J., Gordon, K.H.J., Christian, P.D., 1993. A novel small RNA virus isolated from the cotton bollworm, *Helicoverpa armigera*. *J. Gen. Virol.* 74, 1805–1810.
- Hanzlik, T.N., Dorrian, S.J., Johnson, K.N., Brooks, E.M., Gordon, K.H.J., 1995. Sequence of RNA2 of the *Helicoverpa armigera* stunt virus (*Tetraviridae*) and bacterial expression of its genes. *J. Gen. Virol.* 76 (Pt 4), 799–811.
- Hanzlik, T.N., Christian, P., Gordon, K.H.J., Zeddiam, J.L., 1999. Insect small RNA viruses. *Pest Outlook* 10, 22–26.
- Hanzlik, T.N., Gordon, K.H.J., Gorbalenya, A.E., Hendry, D.A., Pringle, F.M., Ward, V.K., Zeddiam, J.L., 2005. Family *Tetraviridae*. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), “Virus Taxonomy, Eighth report of the International Committee on Taxonomy of Viruses”. Elsevier, Academic Press, Amsterdam, pp. 877–883.
- Helgstrand, C., Munshi, S., Johnson, J.E., Liljas, L., 2004. The refined structure of Nudaurelia capensis omega virus reveals control elements for a T = 4 capsid maturation. *Virology* 318, 192–203.
- Henikoff, S., Henikoff, J.G., 1994. Position-based sequence weights. *J. Mol. Biol.* 243, 574–578.
- Holm, L., Park, J., 2000. DALI: workbench for protein structure comparison. *Bioinformatics* 16, 566–567.
- Huang, X.L., Ali, H., 2007. High sensitivity RNA pseudoknot prediction. *Nucleic Acids Res.* 35, 656–663.
- Johnson, J.E., Munshi, S., Liljas, L., Agrawal, D., Olson, N.H., Reddy, V., Fisher, A., McKinney, B., Schmidt, T., Baker, T.S., 1994. Comparative studies of T = 3 and T = 4 icosahedral RNA insect viruses. *Arch. Virol.* 497–512.
- Johnson, J.E., Reddy, V., 1998. Structural studies of nodaviruses and tetraviruses. In: Miller, L.K., Ball, L.A. (Eds.), “The insect viruses”. Plenum Press, London, pp. 171–223.
- Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227, 680–685.
- Le Gall, O., Christian, P., Fauquet, C.M., King, A.M.Q., Knowles, N.J., Nakashima, N., Stanway, G., Gorbalenya, A.E., 2008. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. *Arch. Virol.* 153, 715–727.
- Liu, X.W., Gorovsky, M.A., 1993. Mapping the 5′ and 3′ ends of Tetrahymena-morphophila messenger-RNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res.* 21, 4954–4960.
- Luke, G.A., de Felipe, P., Lukashov, A., Kallioinen, S.E., Bruno, E.A., Ryan, M.D., 2008. Occurrence, function and evolutionary origins of ‘2A-like’ sequences in virus genomes. *J. Gen. Virol.* 89, 1036–1042.

- Magyar, G., Chung, H.K., Dobos, P., 1998. Conversion of VP1 to VPg in cells infected by infectious pancreatic necrosis virus. *Virology* 245, 142–150.
- Moore, N.F., 1991. The Nudaurelia β family of insect viruses. In: Kurstak, E. (Ed.), "Viruses of Invertebrates". M. Dekker Inc., New York, pp. 277–285.
- Munshi, S., Liljas, L., Cavarelli, J., Bomu, W., McKinney, B., Reddy, V., Johnson, J.E., 1996. The 2.8 angstrom structure of a T = 4 animal virus and its implications for membrane translocation of RNA. *J. Mol. Biol.* 261, 1–10.
- Munshi, S., Liljas, L., Johnson, J.E., 1998. Structure determination of Nudaurelia capensis omega virus. *Acta Crystallogr., D Biol. Crystallogr.* 54, 1295–1305.
- Na, H., Fabian, M.R., White, K.A., 2006. Conformational organization of the 3' untranslated region in the tomato bushy stunt virus genome. *RNA-A Publication of the RNA Society* 12, 2199–2210.
- Nicholas, K.B., Nicholas Jr., N.H.B., Deerfield, D.W., 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNET News* 4, 1–4.
- Olson, N.H., Baker, T.S., Johnson, J.E., Hendry, D.A., 1990. The 3-dimensional structure of frozen-hydrated Nudaurelia-capensis beta-virus, a T = 4 insect virus. *J. Struct. Biol.* 105, 111–122.
- Pan, J., Vakharia, V.N., Tao, Y.J., 2007. The structure of a birnavirus polymerase reveals a distinct active site topology. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7385–7390.
- Pringle, F.M., Gordon, K.H.J., Hanzlik, T.N., Kalkmakoff, J., Scotti, P.D., Ward, V.K., 1999. A novel capsid expression strategy for *Thosea asigna* virus (Tetraviridae). *J. Gen. Virol.* 80, 1855–1863.
- Pringle, F.M., Kalkmakoff, J., Ward, V.K., 2001. Analysis of the capsid processing strategy of *Thosea asigna* virus using baculovirus expression of virus-like particles. *J. Gen. Virol.* 82, 259–266.
- Pringle, F.M., Johnson, K.N., Goodman, C.L., McIntosh, A.H., Ball, L.A., 2003. Providence virus: a new member of the Tetraviridae that infects cultured insect cells. *Virology* 306, 359–370.
- Rambaut, A., Drummond, A.J., Tracer v1.4. (<http://beast.bio.ed.ac.uk/Tracer>). 2007. Ref Type: Computer Program.
- Ranjan, A., Hasnain, S.E., 1995. Influence of codon usage and translation initiation codon context in the Acnpv-based expression system—computer-analysis using homologous and heterologous genes. *Virus Genes* 9, 149–153.
- Reeder, J., Giegerich, R., 2004. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics* 5, 104.
- Reeder, J., Steffen, P., Giegerich, R., 2007. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res.* 35, W320–W324.
- Sabanadzovic, S., Ghanem-Sabanadzovic, N.A., Gorbalenya, A.E., 2009. Permutation of the active site of putative RNA-dependent RNA polymerase in a newly identified species of plant alpha-like virus. *Virology* 394, 1–7.
- Sadreyev, R., Grishin, N.V., 2003. COMPASS: a tool for comparison of multiple protein alignments with assessment of statistical significance. *J. Mol. Biol.* 326, 317–336.
- Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504.
- Schneemann, A., Reddy, V., Johnson, J.E., 1998. The structure and function of nodavirus particles: a paradigm for understanding chemical biology. *Adv. Virus Res.* 50 (50), 381–446.
- Sidorov, I.A., Reshetov, D.A., Gorbalenya, A.E., 2009. SNAD: sequence name annotation-based designer. *Bmc Bioinformatics* 10.
- Strauss, J.H., Strauss, E.G., 1988. Evolution of RNA viruses. *Ann. Rev. Microbiol.* 42, 657–683.
- Swofford, D. L. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. 2000. Sunderland, Massachusetts, Sinauer Associates. Ref Type: Computer Program.
- Taylor, D.J., Johnson, J.E., 2005. Folding and particle assembly are disrupted by single-point mutations near the autocatalytic cleavage site of Nudaurelia capensis omega virus capsid protein. *Protein Sci.* 14, 401–408.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699.
- Xayaphoummine, A., Bucher, T., Isambert, H., 2005. Kinofold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* 33, W605–W610.
- Xu, H.T., Si, W.D., Dobos, P., 2004. Mapping the site of guanylation on VP1, the protein primer for infectious pancreatic necrosis virus RNA synthesis. *Virology* 322, 199–210.
- Yi, F., Zhang, J., Yu, H., Liu, C., Wang, J., Hu, Y., 2005. Isolation and identification of a new tetravirus from *Dendrolimus punctatus* larvae collected from Yunnan Province, China. *J. Gen. Virol.* 86, 789–796.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.