Correspondence H. Bourhy hbourhy@pasteur.fr

# Phylogenetic relationships among rhabdoviruses inferred using the L polymerase gene

H. Bourhy,<sup>1</sup> J. A. Cowley,<sup>2</sup> F. Larrous,<sup>1</sup> E. C. Holmes<sup>3</sup> and P. J. Walker<sup>2,4</sup>

<sup>1</sup>Rabies Laboratory, WHO Collaborating Centre for Reference and Research on Rabies, Institut Pasteur, 28 rue du docteur Roux, 75724 Paris Cedex 15, France

<sup>2</sup>CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia, QLD 4067, Australia

<sup>3</sup>Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA

<sup>4</sup>CSIRO Livestock Industries, Australian Animal Health Laboratory, 5 Portarlington Road, Geelong, VIC 3220, Australia

RNA viruses of the family *Rhabdoviridae* include arthropod-borne agents that infect plants, fish and mammals, and also include a variety of non-vector-borne mammalian viruses. Herein is presented a molecular phylogenetic analysis, the largest undertaken to date, of 56 rhabdoviruses, including 20 viruses which are currently unassigned or assigned as tentative species within the *Rhabdoviridae*. Degenerate primers targeting a region of block III of the L polymerase gene were defined and used for RT-PCR amplification and sequencing. A maximum-likelihood phylogenetic analysis of a 158-residue L polymerase amino acid sequence produced an evolutionary tree containing the six recognized genera of the *Rhabdoviridae* and also enabled us to identify four more monophyletic groups of currently unclassified rhabdoviruses that we refer to as the 'Hart Park', 'Almpiwar', 'Le Dantec' and 'Tibrogargan' groups. The broad phylogenetic relationships among these groups and genera also indicate that the evolutionary history of rhabdoviruses was strongly influenced by mode of transmission, host species (plant, fish or mammal) and vector (orthopteran, homopteran or dipteran).

Received 22 April 2005 Accepted 12 July 2005

# INTRODUCTION

Of the currently described RNA viruses, more than 400 are primarily, though not exclusively, transmitted by arthropod vectors, such as mosquitoes, sandflies, fleas, ticks and lice. These viruses have complex life cycles, with many replicating in both primary and secondary hosts (although the latter may often be dead-ends for transmission) as well as in their arthropod vectors. Before the availability of molecular phylogenetic analysis these viruses were grouped together under the term 'arboviruses', although it is now known that they fall into five phylogenetically distinct viral families, the Togaviridae, Flaviviridae, Bunyaviridae, Reoviridae and Rhabdoviridae. The Rhabdoviridae currently comprises six genera, and members of three of these genera -Vesiculovirus, Lyssavirus and Ephemerovirus - have been obtained from a variety of animal hosts and vectors, including mammals, fish and invertebrates (Tordo et al., 2004). The remaining three rhabdovirus genera are more taxon-specific in their host preference. Novirhabdoviruses infect numerous species of fish, while cytorhabdoviruses and nucleorhabdoviruses are arthropod-borne and infect plants.

All rhabdoviruses contain a single-stranded (-) RNA genome which encodes five virion structural proteins: the

nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the glycoprotein (G) and the polymerase (L) (Dale & Peters, 1981). An added layer of complexity is present in the genus Ephemerovirus, as these viruses contain several additional open reading frames (ORFs) between the G and L genes which encode a second glycoprotein  $(G_{NS})$ and several other non-structural proteins (Walker et al., 1992, 1994; Wang et al., 1994; McWilliam et al., 1997). Similarly, in the genus Novirhabdovirus, a sixth functional cistron between the G and L genes encodes a nonstructural protein (NV) of unknown function (Basurco & Benmansour, 1995). The unclassified rhabdovirus Sigma virus of Drosophila and plant rhabdoviruses in the genera Cytorhabdovirus and Nucleorhabdovirus also contain an additional ORF, which is located between the P and M genes (Heaton et al., 1989; Landes-Devauchelle et al., 1995; Wetzel et al., 1994).

The available gene-sequence data from rhabdoviruses has increased considerably in recent years and this, in conjunction with data on genome organization and a variety of other biological characteristics, has been used for taxonomic classification and species demarcation among the *Vesiculovirus*, *Lyssavirus*, *Ephemerovirus* and *Novirhabdovirus*  genera. In particular, the subdivision of each genus into species is supported by the comparison of nucleotide and deduced amino acid sequences of one (N gene) or several (N and G) common genes (Badrane & Tordo, 2001; Barr *et al.*, 1991; Basurco *et al.*, 1995; Bourhy *et al.*, 1993; Crysler *et al.*, 1990; Kissi *et al.*, 1995; Masters & Banerjee, 1987; Walker *et al.*, 1994; Wang *et al.*, 1995). However, complete genome sequences are available for only a few type species and it is unlikely that such data will be sought for the vast number of unclassified rhabdoviruses: a list of 63 unassigned animal rhabdoviruses is presented in the eighth International Committee on Taxonomy of Viruses (ICTV) report, a further 29 have been only tentatively assigned to genera due to inadequate data (Tordo *et al.*, 2004), and many more are awaiting classification.

One approach to the determination of the phylogenetic relationships among the Rhabdoviridae, as well as the identification of new viral species, is to utilize the conserved amino acid sequence blocks and/or motifs that have been identified in alignments of the RNA-dependent RNA polymerase (L protein) (Bock et al., 2004; Delarue et al., 1990; Dhillon et al., 2000; Elliott et al., 1992; Müller et al., 1994; Le Mercier et al., 1997; Poch et al., 1989, 1990; Tordo et al., 1988; Vieth et al., 2004). Block III of the L polymerase is predicted to be essential for RNA polymerase function because it is conserved among all RNA-dependent RNA polymerases (Delarue et al., 1990; Poch et al., 1989; Xiong & Eickbush, 1990) and mutations in this region abolish polymerase activity (Schnell & Conzelmann, 1995; Sleat & Banerjee, 1993; Jin & Elliott, 1991, 1992). The sequence conservation displayed by this region suggests that it may be a useful target for the exploration of distant evolutionary relationships among the vast array of unclassified rhabdoviruses.

In this study, we inferred the phylogenetic relationships among 56 rhabdoviruses, 20 of which are currently tentative species or unassigned within the *Rhabdoviridae*. This represents the largest phylogenetic study of the *Rhabdoviridae* undertaken to date. Degenerate primers targeting block III of the L gene were defined and used for RT-PCR and sequence analysis, providing a rapid and expansive method to investigate the phylogenetic relationships. The broader goal of this research is to merge phylogenetic and epidemiological information, such as the host and vector species, to provide a more accurate and complete picture of the evolution of key biological characteristics within the *Rhabdoviridae*.

# METHODS

**Virus collection.** A total of 38 rhabdoviruses isolated from various mammalian and insect species were selected from virus collections maintained at CSIRO in Australia and by the Pasteur Institutes in France and Senegal (Table 1). Of these, 15 were unassigned at the species level and five were tentatively classified to a particular rhabdovirus genus. Original virus isolations were predominantly made by intra-cerebral injection in suckling mice (Bourhy & Sureau, 1991)

or passage in C6-36 mosquito cells followed by two to three passages in BHK-21 cells. With many of the viruses, frozen BHK-21 cells or suckling mouse brain stocks were used directly to prepare RNA. Some viruses were additionally grown in 25 mm<sup>2</sup> dishes of BHK-21 cells for 1 to 4 days until a cytopathic effect was recorded.

RT-PCR, cloning and sequencing methods. Total RNA was isolated from virus-infected cells (C6-36 or BHK-21) or virus-infected mouse brain using RNAzol B (Tel-Test Inc.). cDNA was synthesized using 2.5 µg total RNA, 250 ng of each primer and 10 U AMV reverse transcriptase (Promega) in a 20 µl reaction volume using standard methods. Virus L gene fragments were amplified by RT-PCR using the AmpliTaq buffer [2.5 mM MgCl<sub>2</sub>, 200 nM each dNTP, 100 pmol each primer and 2.5 U AmpliTaq DNA polymerase (Perkin Elmer Cetus)] in a 50 µl reaction volume. Amplified DNA was resolved in 2% Ultrapure low-melting-point (LMP) agarose (Gibco-BRL) gels, and DNA products of the expected size were purified using the BresaClean kit (Bresatec). PCR products were either directly sequenced or cloned into the pGEM-T vector and cyclesequenced using universal pUC forward and reverse primers and fluorescent dye terminator FS reagent (ABI). Several independent PCR clones (3-5 clones) were analysed to produce a consensus nucleotide sequence for each virus. Due to genetic variation within the degenerate primer sequences, the primer sequences were excluded from the phylogenetic analysis.

**Phylogenetic analysis.** The dataset of 38 rhabdovirus sequences newly determined here was compared with the corresponding block III L polymerase amino acid sequences of 18 rhabdoviruses collected from GenBank (Table 1). All amino acid sequences were aligned using the CLUSTAL W programme (Thompson *et al.*, 1994) and then checked for accuracy by eye. This resulted in a final alignment of sequences of 158 amino acid residues in length (Fig. 1).

A phylogenetic tree from these data was inferred using the maximumlikelihood method available in TREE-PUZZLE (Schmidt *et al.*, 2002). The WAG model of amino acid substitution was employed along with a gamma ( $\Gamma$ ) distribution of rate heterogeneity among sites, with the value of the shape parameter ( $\alpha = 1$ , for eight rate categories) estimated from the empirical data during tree reconstruction (other parameter values available from the authors on request). Support for each node in the tree was obtained by examining quartet puzzling support values.

# **RESULTS AND DISCUSSION**

# **Design of primers**

On initiation of this study, nucleotide sequence data from the L polymerase gene was only available for 10 members of the genera Lyssavirus, Vesiculovirus and Ephemerovirus. This sequence dataset was first examined to predict degenerate primers that would be broadly reactive among the animal rhabdoviruses. Several degenerate primers were designed to functional L-gene sequences conserved among rhabdoviruses and other negative-sense RNA viruses. No mismatches were tolerated at the 3' end, and less than three mismatches were accepted in the entire sequence, with the exception of Infectious haematopoietic necrosis virus (IHNV). Primers PVO3 [5'-CCADMCBTTTTGYCKYARRCCTTC-3', genome position 7526-7503 in Rabies virus (RV) PV strain] and PVO4 (5'-RAAGGYAGRTTTTYKCDYTR-ATG-3', position 7068-7088), designed to conserved premotif A and to motif B in block III of the L gene, respectively, were found to perform best in preliminary RT-PCR tests

#### Table 1. Isolates of rhabdovirus analysed in this study

UA, Unassigned species and unclassified viruses; TS, tentative species. The abbreviation of the viral name is according to Tordo et al. (2004).

Genus and name	UA/TS	Abbreviation	Reference	Host species/vector	Origin	Year of first isolation	GenBank accession no.
			1101			1001001011	
Nucleorhabdovirus		DVCV	DVCVDI	1. 0			A DO11257
Rice yellow stuntvirus		RYSV	RYSVN	Leathopper			AB011257
Sonchus yellow net virus		SYNV	SYNV	Aphid			L32603
Maize mosaicvirus		MMV	MMV	Leathopper (Peregrinus maidis)			NC 005975
Cytornabaovirus Northern cereal mosaic virus		NCMV	NCMV	Leafhopper (Laodelphax striatellus)	Japan		AB030277
Strawberry crinkle virus		SCV	SCV	Aphid (Fragaria species)	Chile		AY331385
Taastrup virus Novirhabdovirus	UA	TaasV	TV	Leafhopper (Psammotettix alienus)	Denmark		AY423355
Infectious hematopoietic necrosis virus		IHNV	IHNV	Rainbow trout ( <i>Onchorynchus mykiss</i> )/invertebrate reservoirs?			X89213
Viral haemorrhagic septicaemia virus		VHSV	VHSV	Rainbow trout ( <i>Onchorynchus mykiss</i> )/invertebrate reservoirs?			Y18263
Snakehead rhabdovirus	TS	SHV	SR	Snakehead fish ( <i>Ophicephalus</i> striatus)			AF147498
Hiramerhabdovirus	UA	HirR	HR	,			AF104985
Ephemerovirus							
Adelaide Rivervirus		ARV	DPP 61	Bos taurus	Australia	1981	AY854635
Berrimah virus		BRMV	DPP 63	Bos taurus	Australia	1981	AY854636
Kimberley virus	TS	KIMV	CS 368	Bos taurus	Australia	1980	AY854637
Kotonkon virus	UA	KOTV	IbAr23380	Culicoides species	Nigeria	1967	AY854638
Bovine ephemeral		BEFV	BB7721	Bos taurus	?	1968	AY854671
Bovine ephemeral		BEFV	CS 42	Anopheles bancrofti	Australia	1975	AY854639
Bovine ephemeral		BEFV	CS 53	Mixed species	Australia	1974	AY854640
fever virus Bovine ephemeral		BEFV	CS 1933	Bos taurus	Australia	1973	AY854641
fever virus							
Bovine ephemeral fever virus		BEFV	Beijing 1	Bos taurus	China	1976	AY854642
Almpiwar group							
Humpty Doo virus	UA	HDOOV	CS 79	Lasiohelea species	Australia	1975	AY854643
Charleville virus	UA	CHVV	Ch 9824	Phlebotomus species, lizard (Gehyra australis)	Australia	1969	AY854644
Charleville virus	UA	CHVV	Ch 9847	Phlebotomus species		1979	AY854672
Almpiwar virus	UA	ALMV	MRM4059	Ablepharus boutonii virgatus	Australia	1966	AY854645
Oak-Vale virus	UA	OVRV	CS 1342	Culex species	Australia	1981	AY854870
Tibrogargan group							
Tibrogarganvirus	UA	TIBV	CS 132	<i>Culicoides brevitarsis</i> , water buffaloes, cattle	Australia	1976	AY854646
Hart Park group							
Parry Creek virus	UA	PCRV	OR 189	Culex annulirostris	Australia	1972	AY854647
Wongabel virus	UA	WONV	CS 264	Culicoides austropalpalis	Australia	1979	AY854648
Flanders virus	UA	FLANV	61-7484	Culiseta melanura, Culex pipiens quinquefasciatus, Culex salinarus, Culex territans, Culex tarsalis, Seiurus aurocapillus	New York, USA	1961	AF523199
Ngaingan virus	UA	NGAV	NRM14556	<i>Culicoides brevitarsis</i> , wallabies, kangaroos, cattle	Australia	1970	AY854649

#### Table 1. cont.

Genus and name	UA/TS	Abbreviation	Reference no.	Host species/vector	Origin	Year of first isolation	GenBank accession no.
Le Dantec and Kern Canyon group							
Le Dantec virus	UA	LDV	DakHD 763	Human	Senegal	1965	AY854650
Fukuoka virus	UA	FUKV	FUK-11	Culicoides punctatus	Japan	1982	AY854651
Vesiculovirus							
Perinet virus	TS	PERV	Ar Mg 802	Mosquitoes: Anopheles coustani, Culex antennatus, Culex gr. pipiens, Mansonia uniformis; other: Phlebotomus berentensis	Madagascar	1978	AY854652
Vesicular stomatitis New Jersey virus		VSNJV	VSV NJ-H	Sus scrofa/Culex nigripalpus, Culicoides species, Mansonia indubitans	Georgia, USA	1952	AY074803
Vesicular stomatitis New Jersey virus		VSNJV	VSV NJ-O	Bos taurus, equine/Culex nigripalpus, Culicoides species, Mansonia indubitans	Utah, USA	1949	AY074804
Vesicular stomatitis		VSIV	VSV IND	Bos taurus	Indiana, USA	1925	J02428
Indianavirus							110101
Spring viraemia of	TS	SVCV		Cyprinus carpio			018101
Spring viraemia of carp virus	TS	SVCV		Cyprinus carpio	Yugoslavia	1971	AJ318079
Lyssavirus							
Mokola virus		MOKV	Mok	Cat	Zimbabwe	1981	AY854653
Lagos bat virus		LBV	8619NGA	Bat (Eidolon helvum)	Nigeria	1956	AY854654
European bat lyssavirus 1		EBLV-1	8918FRA	Bat (Eptesicus serotinus)	France	1989	AY854655
European bat lyssavirus 1		EBLV-1	9480HOL	Bat (Eptesicus serotinus)	The Netherlands	1987	AY854656
European bat lyssavirus 2		EBLV-2	9337SWI	Bat (Myotis daubentonii)	Switzerland	1993	AY854657
European bat lyssavirus 2		EBLV-2	94112HOL	Bat (Myotis dasycneme)	The Netherlands	1989	AY854658
Duvenhage virus		DUVV	94286SA	Human	South Africa	1986	AY854659
Australian bat lyssavirus		ABLV	ABLh	Human	Australia	1998	AF418014
Australian bat lyssavirus		ABLV	ABLb	Bat (Pteropus species)	Australia	1996	AF081020
Rabies virus		RABV	9911	Dog	Cambodia	1998	AY854660
Rabies virus		RABV	PV	Vaccine			AY854661
Rabies virus		RABV	SADB19	Vaccine			AY854662
Rabies virus		RABV	9706	Vaccine AG	China		AY854663
Rabies virus		RABV	8743	Human	Thailand	1983	AY854664
Rabies virus		RABV	9702	Human	India	1997	AY854665
Rabies virus		RABV	9106	Human	Morocco	1990	AY854666
Rabies virus		RABV	9704	Bat (Tadarida brasiliensis)	Argentina	1997	AY854667
Rabies virus		RABV	9147	Fox	France	1991	AY854668
Rabies virus		RABV	02008	Bat	USA	1994	AY854669

with selected viruses and were thus used in subsequent tests. Amino acid and nucleotide sequence alignments from which the primer sequences were designed are shown in Fig. 1.

Primers PVO3 and PVO4 produced PCR products for 38 animal rhabdoviruses, 20 of which are not currently assigned to a particular genus (Table 1). These primers amplified a 456–462 nucleotide region, conforming to the expected size. Internal primers PVO5 (5'-ATGACGG-ACAAYCTGAACAA-3', position 7170–7189) and PVO6 (5'-CCRTTCCARCAGGTAGGDCC-3', position 7486–7467) were used for sequencing some PCR products. These primers amplified a 317 nucleotide region.

#### Sequence analysis of L polymerase block III

A total of 56 rhabdovirus L polymerase sequences were subjected to phylogenetic analysis (Table 1). These sequences encompass the three highly conserved segments (pre-motif A, motif A and motif B) of block III of the L polymerase (Fig. 2), which is present in all the



**Fig. 1.** Rationale for the design of primers. Conserved amino acid (a) and nucleotide sequences (b) in the L genes of RV strains PV (Genbank accession no. M13215) and SAB-B19 (M31046), VSV serotypes Indiana (K02378) and New Jersey, strains Ogden (M29788) and Hazelhurst (AY074803), bovine ephemeral fever virus (BEFV) (AF234533), IHNV (X89213) and *Sonchus yellow net virus* (SYNV) (M87829) to which degenerate PCR primers PV03 and PV04 were targeted. The 3' termini of primers PV03 (384-fold degenerate) and PV04 (576-fold degenerate) were designed to invariant Met and Glu residues, respectively, and amino acids conserved in prototype viruses of at least four of the five genera compared are shaded. Amino acid positions are indicated according to the polymerase sequence of RV PV (Tordo *et al.*, 1988).

RNA-dependent RNA polymerases studied so far, including reverse transcriptase (Poch *et al.*, 1989; Xiong & Eickbush, 1990). Although these sequences are extremely divergent, sufficient sequence similarity exists in some domains of the rhabdovirus polymerases to make phylogenetic analysis possible (Dhillon *et al.*, 2000; Le Mercier *et al.*, 1997; Müller *et al.*, 1994; Vieth *et al.*, 2004). Importantly, the alignment confirmed the conservation of some residues among all the *Rhabdoviridae* (Le Mercier *et al.*, 1997; Müller *et al.*, 1994), whilst also identifying new residues that are conserved among the Mononegavirales (Fig. 2).

Structural conservation of residues in the aligned sequences was examined by using similarities in charge or polarity and matching aromatic residues (Poch *et al.*, 1990) as the primary criteria to define the following amino acid families: [P, G; S, T; A] [F, Y, W; I, L, M, V] [D, E; N, Q] [K, R, H] [C]. Only residues belonging to the same family and conserved in at least 54 of the 56 sequences and among all genera were considered as meaningful indicators of sequence homology. Ninety-nine amino acid residues were conserved among the entire *Rhabdoviridae* family (Fig. 2). Among these, 12 were conserved among the entire *Rhabdoviridae* and the *Paramyxovirinae* subfamily. Finally, five positions were conserved among the Mononegavirales and all known pre-motif A and motif A (Dhillon *et al.*, 2000; Le Mercier *et al.*, 1997;

Müller *et al.*, 1994; Vieth *et al.*, 2004): R (562), E (569), D (618), K (621) and F/I (648).

# Phylogenetic analysis of the *Rhabdoviridae* using the sequence of block III

Previously, taxonomic relationships among members of the Rhabdoviridae were primarily based on structural properties (genome size and complexity), large-scale biological properties (host range, epidemiological cycles, routes of transmission) and serological cross-reactions (immunofluorescence, complement fixation (CF), neutralization tests). Although serological data are useful taxonomic tools for closely related viruses, their interpretation in defining relationships among more distantly related viruses has proven complex (Calisher et al., 1989; Shope, 1995; Wang et al., 1995). More recently, the extent of sequence similarity within a given gene has largely been used for species demarcation in each genus of the Rhabdoviridae. In the Lyssavirus genus, for instance, percentage sequence similarity within the nucleoprotein gene has been used for the definition of different virus genotypes (Arai et al., 2003; Bourhy et al., 1993; Kuzmin et al., 2003), and the same methodology has been used for the delineation of different species among the vesiculoviruses and ephemeroviruses (Barr et al., 1991; Crysler et al., 1990; Masters & Banerjee, 1987; Walker et al., 1994; Wang et al., 1995).

558								609
!-P	re Motif	A	- !					[ -
SM	MURLYFVT	TEKI	LANYTLP	LEDALTI	MTDNI NKV	FKKLTI	DRVTGO	GLLD-YS
CIM	NUDLVEVI	TEKI	LANVILD		ATTONI NKV	FRELTI		QLLD-VG
Chr.		TERL	I ANYTI D	LEDAL II	ALL DIVIDIVICY.	FICILIE TI		GLLD VC
50	NLRLIFVI	TERL	JUANIILP.	LFDALTI	TTDNLINKV.	FRALL	DRVIGQ	GLLD-15
SW	NTKTAFAT	TEKI	LANYILP	LFDALTI	MTDNLNKV.	FKKLII	DRVTGQ	GLLD-YS
SW	NLRLYFVI	TEKI	LANYILP	LFDALTI	MTDNLNKV	FKKLII	DRVTGQ	GLLD-YS
SW	NLRLYFVI	TEKI	LANYILP	LFDALTI	MTDNLNKV.	FKKLII	DRVTGQ	GLLD-YS
SW	NLRLYFVI	TEKI	LANYILP	LFDALTI	MTDNLNKV.	FKKLII	DRVTGQ	GLLD-YS
SW	NLRLYFVI	TEKI	LANYILP	LFDALTI	ATDNLNKV	FKKLII	DRVTGO	GLLD-YS
SM	NURLYEVT	TERI	<b>LANYTLP</b>	LEDALTI	MTDNLNKV	FKKLTI	DRVTGÕ	GLSD-YS
SM	NLRLYFVT	TEKI	TANYTT.P		MTDNI NKV	FKKLTI		GLLD-VS
Chi	NE DI VEVI	TERI	LANVIID		AUTONIC METZ	FRETTI		CLOD VC
OW.	NUCLIFVI	TERI	JUANTI DP		ADDMLMRV.	PKKI TI	DRVIGQ	GLOD NG
SW	NLRLIFVI	TERI	LANTLP	LFDALTI	MITDIALDIK V.	PKKLII	DRVTGQ	GLQD-YS
SW	NTKTAFAT	TEKI	LANYILP	LFDALTI	MTDNLNKV.	FKKTTI	DRVTGQ	GLQD-YS
SW	NLRLYFVI	TEKI	LANYILP	LFDALTI	MTDNLNKV	FKKLII	DRVTGQ	GLQD-YS
SW	NLRLYFVI	TEKI	LATYILP	LFDALTI	MTDNLNKV.	FKKLII	DRVTGQ	GLLD-YS
SW	NLSLYFVI	TEKL	LATYILP	LFDALTI	MTDNLNKV.	FKKLII	DRVTGQ	GLLD-YS
SW	NLRLYFVI	TEKL	LATYILP	LFDSLTI	TONLNKV.	FTKLI	DRVTGO	GLLD-YS
SM	NURLYFVT	TEKI	LANHITP	LEDALTI	MTDNLNKV	FKKLTI	DRVTGO	GLKD-YS
SM	MLRLVFVT	TEKI	LANHTLP		MTDNL NKV	FKKLTI	DRVTGO	GLRD-VS
Cbl	FI DEVEVE	TEVI	TEVEVEND	ם ביים ביים וייי דיים ד	ADDIOGU	TEEMI	ENCOCO	CADD-VE
0.0	EDREIFVF	TETT	TREVEND	LENQTI	ADDDQ3V	TREAT	ENSQGQ	GOND VE
50	ELKEIFVF	J.F.XI	JIKEIFVP	LFHGLT	ADDLQSV	TKKMP	ENSQGQ	GSND-YE
SW	ELREYFVY	TEAT	TKEJEAD	LFHGLTI	MADDLQEV	TKKMP	ENVSGQ	GLDN-YE
TW	ELREYFVC	TEYN	IIKQFFIP	LFQGLTI	MADDMQEV	LKKLL;	SSSSGQ	GLDN-YN
SW	QLREYFVI	TEYI	JIKTHYVP	LFKGLTI	MADDLTSV	VKKMLI	DNTNGQ	GLDD-YS
SW	QLREYFVI	TEYI	JIKTHYVP	LFKGLTI	MADDLTSV	VKKMLI	DNTNGQ	GLDD-YS
SW	RLREYFVI	TEYI	JIKTYYVP	LFKGLTI	MADDLTSV	IKKMMI	DSSSGQ	GLDD-YS
SW	RLREYFVI	TEYI	JIKTYYVP	LFKGLTI	MADDLTSV	IKKMMI	DSSSGQ	GLDD-YS
SW	KFPEYFVI	TEYL	IKTHFVP	MFKGLTI	MADDLTAV	IKKMLI	DSSSGQ	GLKS-YE
SW	RLREYFVI	TEYL	IKTHFVP	LFHGLTI	MADDMTAV	ткким	ESSSGO	GLND-YS
SY	ELRDYFVS	TEYL	JIKKYFVP	LFEGLTI	ADDLNTV	IKKMLJ	DVSSGÕ	GTRE-YE
SY	ELEDYEVS	TEYI	TKKYFVP	LEEGLTI	VTRADDLATV	тккмы	DVSSGO-	GTRE-YE
SV	ELEDYEVS	TEVI	TKKYFVP	LFEGLTI	ADDLNTV	TKKML	DVSSG0	GTRE-VE
ev	ELIDIT VO	TEVI	TEEVEVD	T RROT M		TEEMIT		VE
ev	EDRETT VO	TETT	TERRITUR	T RECT T		TVVMII		CTPE-VE
01	ELKDIF VG	TEL	TREAT	LEGUI	ADDINIV	TREAT	DV35GQ	OWDE VD
51	ELRDIFVS	TET	JIKKIFVP	LFEGLI	ADDINIV		DVSSGQ	GREATE AND
SY	ELRDYFVA	TEYL	JIKKYYVP	LFEGLTI	MADDLNTV	IKKMLI	DVSSGQ-	GTRN-YD
SW	ETKEALAL	TEAL	TKEHFVP	LFKGLTI	MADDLQTV	TKKWTI	DVSAGQ	GTET-YE
SY	DMRDYFVM	TEYL	JIKKYYVP	LFKGLTI	MADDQNTV	VKKMLI	KVSKGQ	GLTD-YK
TW.	NLRNYFVM	TELL	IKEHFIG	LFNGLTI	MADDLQGL	IKKLL	DRTTGQ	GDSK-IK
SW	KLREYFVI	TEWI	IKHHFIP	LFEGLTI	MADDMNTV	ITKMII	NKTSGQ	DKKSCED
SW	KLREYFVI	TEWI	IKHHFIP	LFEGLTI	ADDMNTV	ITKMII	NKTSGQ	DKKSCED
SW	KLREYFVI	TEWL	IKHHFIP	LFEGLTI	MADDMNTV	ITKMI	NKTSGQ	DKKSCED
SW	SLREYFVV	TEFL	IKKHFLP	LFDĞLTI	SDDLNTV	ITKMI	GKTIGO	DEHSSOK
SW	ALREYEVT	TEYI	TKTHYVP	LESGLTI	VTTJUCAN	ISKLIJ	$DRTOGO - \cdot$	GGID-YD
TW	FLREVEVI	TEVI	TKTHEVD	LESCLT		LORULI	DRTOGO	VE
CIN		TETT	TUEUETD			TCKMT	epenco	CPDD-VK
		TELL	U CENTLD	NIDGIU	ADDDDDDV	I GIULI	SKSDGQ	CKOCE E
TL	NMRSIVVI I I DUMMAN	TEMP	LODUTLY	VEDOTE	MINMEDE.	NUCLEAR	RATESQ	GRQGEF
SH	LLRVYVVL	TEQI	LSDHILK	TEPQIT	MUDULEDD.	TKKPIT	STVRHQS	ALINKKRGSDRTWA
TF	EKRTYVVL	TESI	JAEYILP	LFPEIT	MMDDEIKL	LKKMF:	SATNNS	QGQ
SF	RLRLYCVS	TEAL	LGDKILK	AEBÖTTI	ISLDMLTM	TKKWFI	KVSSQ1'I'	REDD
SF	KMRLYFTA	TEEI	JGSKLLR	YFPQITI	ASSNLLDM	QEKMS:	SMSRDLE	SQNK
SY	KLRMYVTŠ	TEEI	'TCKAATK	YFPMITI	ISDNLLSM	VIRLFI	DMTTLIG	DKGV
TF	MPRLLQVL	RES-	IAKKTSK	LFPEITI	MTSSDLDM	KKRKFI	MLSKRS	DDRRG
TF	MPRLLQVL	RES-	IAKKTQK	LFPEITI	MTSSDLDM	KKRKFI	MLSKKS	DDRRG
TF	KPRLLQVL	RES-	IAKKTSK	LFPEITI	TASDLDL	KKRKFI	LVSRKS	DDRRG
AF	RPRLLQVL	RES-	IAKKTSK	LFPEITI	MTFSDLEL	KKKKF	QLSRKS	DDRRG
.SW	R	Е	I	F ITI	n v	K		
TY	_	_	L	Ŀ	м	R		
AF				-	L			

Our phylogenetic analysis of the 158-residue L polymerase sequence produced an evolutionary tree that generally, although not entirely, conformed to accepted serological groupings and taxa within the Rhabdoviridae (Calisher et al., 1989; Shope, 1995; Tordo et al., 2004). In particular, members of four genera - Lyssavirus, Novirhabdovirus, Cytorhabdovirus and Nucleorhabdovirus - obtained from a variety of host species, including mammals, fish, arthropods and plants, can be easily distinguished and fall into relatively well-supported clades (Fig. 3). Although the vesiculoviruses and ephemeroviruses also fell into clear monophyletic groups, they are less well supported by quartet puzzling, and each genus contained some unclassified viruses. Furthermore, Kotonkon virus, which causes clinical ephemeral fever in cattle (Kemp et al., 1973; Tomori et al., 1974), but which has previously been classified as a lyssavirus, very clearly clustered with members of the genus Ephemerovirus. Lastly, there is some evidence that the two groups of plant rhabdoviruses - the cytorhabdoviruses and nucleorhabdoviruses - form a distinct clade, although this

610			669
!			
KALIVLUDI EKMININDEL EGLEI	VESVLDQ	FGLKKVF;	SKINEFFQKSWIIISDKSDLI
RVTVAEHLDVEKMNNHORLESTEL	VESVLOO	JEGLKEVE,	SRTHEFFORSWITTSDRSDLI
RVTYAFHLDYEKWNNHORLESTEL	VESVLDO	FGLKEVES	SETHEFFORSWITTSDRSDBI
RVTYAFHLDYEKWNNHORLESTEL	VFSVLDOV	FGLKEVE	SRTHEFFORSWIYYSDRSDLI
RVTYAFHLDYEKWNNHORLESTEI	VFSVLDO	/FGLKRVF	SRTHEFFOKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHORLESTEI	VFSVLDO	/FGLKRVF:	SRTHEFFOKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTEI	VFSVLDQ	/FGLKRVF:	SRTHEFFQKAWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTEI	VFSVLDQ	VFGLKRVF:	SRTHEFFQKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTEI	VFSVLDQ	JFGLKRVF:	SRTHEFFOKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDY	JFGLKKVF:	SRTHEFFQKSWVYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDY	JFGLKKVF:	SRTHEFFQKSWVYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDQ	JFGLKKVF	SRTHEFFQKSWVYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFPVLDQV	JFGLKKVF:	SRTHEFFQKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDKV	/FGLKKVF:	SRTHEFFQKSWVYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDK	/FGLKKVF:	SRTHEFFQKSWVYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTII	VFSVLDK	JFGLKKVF:	SRTHEFFQKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDRA	AFGMKKVF:	SRTHEFFQKSWIYYSDRSDLI
RVTYAFHLDYEKWNNHQRLESTKI	VFSVLDK	AFGLSHVF:	SRTHEFFQKSWIYYSDRSDLI
FVSIANHIDYEKWNNHQRKESNYY	VFRVMGQQ	CFGLPNLF'	TRTHEFFEKSLIYYPQRADLM
YVSIANHIDYEKWNNHQRKESNYY	VFKVMGQG	CFGLPNLF'	TRTHEFFEQSLIYYPQRADLM
YISIANHIDYEKWNNHQRYESNCH	IIFKVMGQG	CFGLPNLFI	LRSHEFFQKSLIYYNQRPDLM
LISIANHFDYEKWNNHQRYASNCH	IVFKVMGQG	JFGLPNLFI	LRTHEFFEKSLIYYANRPDLM
SICIANHIDYEKWNNHORKESNGE	VERVMGQI	TLGYPRLFI	ERTHEFFESSLIYYNGRPDLM
SICIANHIDYEKWINHQRKESNGE	VERVMGQI	FLGYPRLFI	ERTHEFFESSLIYINGRPULM
SVCLANTIDI EQWININQRKESINGE		ZLGIPSLII ZLOVDČI II	CRINEPPERSITYINGEPULL
A TOTANUT DVERMINNOPRI SNOT		ZLGIPSLII ZLOVDOLTI	CRINEPPERSETTINGRPDEM
STCLANHIDIEKWNNHORKESNGE	VERVMCO	PLORDNLTS	28THEFFER CLIVVNCPDDIM
VITIANNI DVEKMNNOR LESNOE	VETVIGQ	FLGI DNI F	PRTHEFFORSI I VVNORDDI M
VITIANNIDIEKWNNIQKIESNO	VETVMCO	FLGLOMLE	PRTHEFEORSI. IVVNOR DDI.M
YTTIANNI DYEKWNNYOR IESNOF	VETVMGO	FLGLPNLF'	TRTHEFFORSLIYYNORPDLM
YTTIANNIDYEKWNNYORIESNGE	VETVMGRI	FLGLPNLF'	TRTHEFFORSLIYYNORPDLM
YITIANNIDYEKWNNYORIESNGE	VFTVMGO	LGLPNLF	TRTHEFFORSLIYYNORPDLM
YITIANNIDYEKWNNYÖRIESNGE	VFTVMGŘI	FLGLPNLF	TRTHEFFOKSLIYYNORPDLM
YVSIANNIDYEKWNNYQRKESNGE	VFRVMGQI	LGMENLI	/RTHEFFENSLVYYNQRADLM
NITIANNIDYEKWNNYQRYDSNSA	IFTVMGQ	FLGYPKLIA	ARTHEFFEKSLIYYNQRPDLM
EITFANHLDYEKWNNYQRRESNGE	VFRVMGQI	FLGLPHLI	RTHEFFENSLIYYNGRPDLM
KINIANGLDYTKWNNYQRYDSNRY	VFRVMGQI	FLGYDMLI	ERTHQFFEQSLIYYPQRPDLM
IITICNHLDYEKWNNNQRGASNNE	VFRVMGQI	FGYPRLI	ERTHEIFENSFIYFVNRPDLM
IITICNHLDYEKWNNNQRGASNNE	VFRVMGQI	FFGYPRLI	ERTHEIFENSFIYFVNRPDLM
IITICNHLDYEKWNNNQRGASNNE	VFRVMGQI	FLGYPRLI	ERTHEIFENSFIYFVNRPDLM
IVTISNHLDYEKWNNNQRAESNDE	VFKVMGQ3	FLGYPNLI	TRTHEIFQKSLIYFVNRPDLM
NICIANHIDYEKWNNHQRLESTGE	VFKVMGQ3	FLGYPNLI	WRTHEFFEKSLVYYNGRPDLM
NICIANHIDYEKWNNHKRFESTRY	VFKVMGQI	FLGYPSLII	EITHLIFQKCFVYFTDRPDLM
EVTYANHMDYSKWNNHQRGKINNE	PTFKVMGM	FLGYPKLII	ERTHEIFEKSLIYYAGDKTLL
SRTFCINMDFEKWNLNMRKEGTYY	VFQELGRI	LFGLPTLYI	NKTYDIFRNSTIYLADGSYNP
SKVICMSLDFEKWNGHMRKEMTLG	SVFTPIGDI	LFGMTELYI	NVTYDIFSECYYYLADGTYVP
EKSIIVSFDFMKWNSNMRFEETTU	VECOLONI	LEGENNCII	NRTHUMFNEGIIYLALGTYVP
SVIVIFNLDFIKWNLQMRRNICEF	VESQUGKI	LECT DOL V	RIHETFRUSLITLCSGEGVL
AVTY CMNT DE CRUNOM DE DENNA	F LENNSEI	LICEPCIT:	SOURCENSON ALCOCOAND
ETHINKSLDINKECTSOROENSS	WESSLDE	MCTEPLE'	SRVHEIFEKTWIVDGSASDPP
FIHVNKSLDINKFCTSOROFNSNA	WESSLDE	MGTEPLES	SEVHETEEKTWIVDGSSSDPP
YVHMSKSLDINKFCTSOROFNSOF	VFOCLDEI	LLGTGALF	SRVHEIFEKTWIVDGSASDPP
YIHISKSLDINKFCTSOROFNSLA	VFOSLDEI	LLGTDOLF	TRVHEIFEKTWIVDGSASDPP
ID KW R	VF	FG LF	R F IY
L F	IL	L VI	LV
М	F	М	V
F	т		

has relatively low quartet puzzling support. Taastrup virus, which was unassigned (Bock et al., 2004), is related to cytorhabdoviruses.

Strikingly, our phylogenetic analysis also identified four more monophyletic groups of currently unclassified rhabdoviruses which have variable support values. First, Wongabel, Parry Creek, Flanders and Ngaingan viruses formed a distinct cluster, with high levels (93%) of quartet puzzling support. We refer to this as the 'Hart Park' group, based on the serologic grouping of Flanders virus in the Hart Park serological group (Boyd, 1972; Calisher et al., 1989). Second, a tentatively named 'Almpiwar' group, containing Almpiwar virus, Humpty Doo virus, Charleville virus and Oak-Vale virus, was also identified. Although this grouping had only 65 % quartet puzzling support, Almpiwar virus and Charleville virus possessed almost indistinguishable sequences in the L-gene region, and both have been associated with infection in lizards. Another group, consisting of the Le Dantec and Fukuoka viruses, and herein referred

|--|

!=-Motif B    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    REDQIYCLDMSNGPTCWNGQDGGL    REDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSDGPTCWNGQDGGL    WEDQIYCLDMSDGPTCWNGQDGGL    WEDQIYCLDMSDGPTCWNGQDGGL    WEDQIYCLDMSDGPTCWNGQDGGL    WEDQIYCLDMSDGPTCWNGQDGGL    WEDQIYCLDMSDGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQDGGL    WEDQIYCLDMSNGPTCWNGQEGGL    WEDQIYCLDMSNGPTCWNGQEGGL    WEDQIYCLDMSNGPTCWNGQEGGL    WEDQIYCLDMTNGPTCWNGQEGGL    WEDQIYCLDMTNGPTCWNGQEGGL    WEDQIYCLDMT	70		695
MEDQ I I LDMS - NGPT CNNG - QDGQL MEDQ I Y CLDMS - DGPT CNNG - QDGGL MEDQ I Y CLDMS - DGPT CNNG - QDGGL MEDQ I Y CLDMS - DGPT CNNG - QDGGL MEDQ I Y CLDMS - NGPT CNNG - QDGGL (MEDQ I Y CLDMS - NGPT CNNG - QDGGL (MEDQ I Y CLDMS - NGPT CNNG - QDGGL (MEDQ I Y CLDMS - NGPT CNNG - QGGGL (MEDQ I Y CLDMS - NGPT CNNG - QKGGL (MEDQ I Y CLDMS - VI V CWEG - QKGGL (MEDQ I Y CLDMS - VI V CWEG - QKGGL (MEDSL VNTTD KNV CWEG - QKGGL (MGRE CLNRLG - VKV CWEG - QKGGL (MGR	OF WEDOTVOLDMO	! <u>Moti</u>	<u>f B</u> !
MEDQ I L LDNS - NOFTONNG QDGQL WEDQ I YCLDMS - NGPTONNG QDGGL WEDQ I YCLDMS - NGPTONNG QDGGL WEDQ I YCLDMS - NGPTONNG QDGGL WEDQ I YCLDMS - DGPTONNG QDGGL WEDQ I YCLDMS - NGPTONNG QDGGL WEDQ I YCLMS - NGPTONNG QDGG WEDQ I YCLMS - NGPTONNG QGGG WEQGSL WNNS CWVCWEG - QKGGL RGBSLWNTD KWVCWEG - QKGGL RGBSLWNSN CWVCWEG - QKGGL RGRECLNRLG - VKVCWEG - QKGGL CGRHCLNSKG - VVCVEG - QKGGL CGCH CLNSKG VKVCWEG	GEWEDQIICEDMS	INGPICWING===	ODGGI
MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQPGQL REDQ1YCLDMSNGPTCWNGQPGQL REDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQDGQL MEDQ1YCLDMSNGPTCWNGQAGQL MENQ1YCLDMAAGPTCWNGQKGQL MENQ1YCLDMAKIVCWEGQAGGL MENQ1YCLDMAKIVCWEGQAGGL MENQ1YCLDMAKIVCWEGQAGGL MENQ1YCLDMAQVCWNGQKGQL MENQ1YCLDMAKIVCWEGQAGGL MENQ1YCLDMA	GLWEDQIYCLDMS	NGPTCWNG	QDGGL
MEDQ1YLLDMS    NGPTCWNG    QDGQL      MEPQ1YLLDMS    NGPTCWNG    QYGQL      MEPQ1YLLDMS    NGPTCWNG    QYGQL      REDQ1YLLDAS    NGPTCWNG    QDGQL      REDQ1YLLDAS    NGPTCWNG    QDGQL      MEDQ1YLLDAS    NGPTCWNG    QDGQL      MEDQ1YLLDMS    DGPTCWNG    QDGQL      MEDQ1YLLDMS    NGPTCWNG    QDGQL      MEDQ1YLLDMS    NGPTCWNG    QDGQL      MENQ1YLLDMS    NGPTCWNG    QDGQL      MENQ1YLLDMS    NGPTCWNG    QDGQL      MEQQ1YLLDMS    NGPTCWNG    QDGQL      MEQQ1YLLDMS    NGPTCWNG    QDGQL      MEQQ1YLLDMS    NGPTCWNG    QDGQL      MEQQ1YLLMSA    HNCYCWGG    QGGGL      MEGNTLWNSA    NLVCWNGG    QAGGL	GLWEDQIYCLDMS	NGPTCWNG	QDGGL
MEDQ 11 CLDMS	GLWEDQIYCLDMS	NGPTCWNG	QDGGL
MEDQ11CLDBS    NOPTCWNG    QFGG1      REEDQ1YCLDAS    NOPTCWNG    QDGGL      REEDQ1YCLDAS    NOPTCWNG    QDGGL      MEDQ1YCLDMS    DGPTCWNG    QDGGL      MRDQ1YCLDMS    NGPTCWNG    QDGGL      MRDQ1YCLDMS    NGPTCWNG    QDGGL      MRDQ1YCLDMS    NGPTCWNG    QDGGL      MRDQ1YCLDMA    AGPTCWNG    QDGGL      MRDQ1YCLDMA    AGPTCWNG    QDGGL      MRDQ1YCLDMA    AGPTCWNG    QCGGL      MRDQ1YCLDMA    AGPTCWNG    QKGGL      MCGSLUNNSP    -    YLVCWGG    QKGGL      MCGDSLUNTP    -    KWCWEG    QKGGL      MGBSLUNTTP    -    KWCWGG    QKGGL      MGBSLUNTTP    -    KWCWEG    QKGGL      MGRDSLUNTTP	GLWEDQIYCLDMS	NGPTCWNG	QDGGL
AREDQ1YCLDASNAPTCWNGQDGQL MEDQ1YCLDASNAPTCWNGQDGQL MEDQ1YCLDASNAPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSDGPTCWNGQDGQL MEDQ1YCLDMSNAPTCWNGQDGQL MEDQ1YCLDMSNAPTCWNGQDGQL MEDQ1YCLDMSNAPTCWNGQDGQL MEDQ1YCLDMSNAPTCWNGQRGQL MEDQ1YCLDMSNAPTCWNGQRGQL MEDQ1YCLDMSSLVCWNGQRGQL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSASLVCWNGQKGGL PGGDTLTNSSAQKVCWEGQKGGL PGGDTLTNSSAQKVCWEGQKGGL PGGDTLCNSTKHRVCWNGQKGGL PGGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGRECLNRLGVKVCWEGQKGGL PGGRECLNRLG	GLWEDQIYCLDMS==-	NGPTCWNG	QIGGL
NAEDQ1YCLDAS NGPTCWNG    ODG0L      WEDQ1YCLDMS	GLREDQIYCLDAS	NGPTCWNG	QDGGL
MEDQ1YCLDJSNGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMTNGPTCWNGQDGQL      MEDQ1YCLDMTNGPTCWNGQDGQL      MEDQ1YCLDMTNGPTCWNGQDGQL      MEDQ1YCLDMTNGPTCWNGQDGQL      EWRQ1YCLDMSMGPTCWNGQDGQL      EWRQ1YCLDMSMGPTCWNGQDGQL      EWRQ1YCLDMSNGPTCWNGQDGQL      EWRQ1YCLDMS	GLREDQIYCLDAS	NGPTCWNG	QDGGL
MEDQ1YCLDBSDAPTCWNGQDGGL      MEDQ1YCLDMSDGPTCWNGQDGGL      MEDQ1YCLDMSDGPTCWNGQDGGL      MEDQ1YCLDMSDGPTCWNGQDGGL      MEDQ1YCLDMSDGPTCWNGQDGGL      MRDQ1YCLDMTNCPTCWNGQDGGL      MRDQ1YCLDMTNCPTCWNGQDGGL      MRDQ1YCLDMTNCPTCWNGQDGGL      MRDQ1YCLDMTNCPTCWNGQDGGL      MRDQ1YCLDMTNCPTCWNGQDGGL      MRDQ1YCLDMTSLVCWNGQKGGL      VEGDTLINNSASLVCWNGQKGGL      VEGDTLINNSASLVCWNGQKGGL      VEGDTLINNSASLVCWNGQKGGL      VEGDTLINNSASLVCWNGQKGGL      VGGBLINNTEHNVCWEGQKGGL      VGBSLWNTDKIVCWEGQKGGL      VRGBLVNNSKRNVCWEGQKGGL      VRGBSLWNTDKIVCWEGQKGGL      VRGBSLWNSKREVCWNGQKGGL      VGRSLUNNSN	GLWEDQIYCLDIS	NGPTCWNG	QDGGL
MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MRDQ1YCLDMSNGPTCWNGQDGQL      MRDQ1YCLDMSNGPTCWNGQDGQL      MRDQ1YCLDMSNGPTCWNGQDGQL      MRDQ1YCLDMSNGPTCWNGQDGQL      MRDQ1YCLDMSNGPTCWNGQDGQL      MRDQ1YCLDMASGPTCWNGQDGQL      MENDQ1YCLDMASGPTCWNGQDGQL      MENDQ1YCLDMASGPTCWNGQDGQL      VEGNTLVNNSPYLVCWDGQKGQL      PEGDTLTMSSASLVCWNGQKGQL      MEDQ1YCLDMTKNVCWEGQKGQL      MEGDSLVNTTDKNVCWEGQKGQL      CRGDSLVNTTDKNVCWEGQKGQL      CRGDSLVNTTD	GLWEDQIYCLDMS	DGPTCWNG	QDGGL
MEDQ1YCLDMS DGPTCWNG QDGCL      WEDQ1YCLDMS DGPTCWNG QDGCL      WEDQ1YCLDMS DGPTCWNG QDGCL      WRDQ1YCLDMT NGPTCWNG QDGCL      WRDQ1YCLDMT SLVCWNG QKGCL      /EGNTLWNSP YLVCWDG QKGCL      /EGNTLWNSP YLVCWNG QKGCL      /EGNTLWNSP YLVCWNG QKGCL      /RGD5LWNTD KLVCWRG QKGGL      RGBSLWNTD KLVCWNG QKGGL      RGRDSLWNST HRVCWNG	GLWEDQ1YCLDMS	DGPTCWNG	QDGGL
MEDQ1YCLDMSDGPTCWNGQDGQL      MEDQ1YCLDMSDGPTCWNGQDGQL      MRDQ1YCLDMTNGPTCWNGQDGQL      MRDQ1YCLDMTNGPTCWNGQDGQL      MRDQ1YCLDMTNGPTCWNGQDGQL      WRDQ1YCLDMTSCPTCWNGQDGQL      WRDQ1YCLDMASCPTCWNGQDGQL      WRDQ1YCLDMASCWNGQCGC      JEDRLESIDPNYLVCWNGQKGGL      JEDRLESIDPNQLVCWNGKLEDCRMRQKGWS      YGGD1LWNSPKIVCWEGQAGGL      JRGDSLWNTDKIVCWEGQAGGL      JRGBSLWNTDKIVCWEGQAGGL      JRGBSLWNTDKIVCWEGQAGGL      JRGBSLWNTDKIVCWEGQAGGL      JRGBSLWNTDKIVCWEGQKGGL      JGGNSLWNSNQPVCWQGQKGGL      JGGNSLWNSNQPVCWQGQKGGL      JGGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JRGRECLNRLGVKVCWEGQKGGL      JQNNTLQNRGNSLVCWQGQLGGL      JQNNTLQNRGNSLVCWQGQLGGL      JQNNTLQNRGNSLVCWQGQLGGL      JDPUSUFFEDECRVCWQGQLGGL	GLWEDQIYCLDMS	DGPTCWNG	QDGGL
MEDQIYCLDMS DQPTCVNG QDGGL MRDQIYCLDMT NGPTCVNG QDGGL MRDQIYCLDMT NGPTCVNG QDGGL MRDQIYCLDMS NGPTCVNG QDGGL WRDQIYCLDMA AGPTCVNG QDGGL WRDQIYCLDMA AGPTCVNG QGGL PLEDELESTDPN QLVCVNGKLEDGRMRQKGWS QGCBLINNTE	GLWEDQIYCLDMS	DGPTCWNG	QDGGL
MRDQIYCLDMT NGPTCWNG QDGGL WRDQIYCLDMT NGPTCWNG QDGGL WRDQIYCLDMT GPTCWNG QDGGL WRDQIYCLDMT GPTCWNG QDGGL WRDQIYCLDMT GPTCWNG QRGGL PGNTLVWNSP TLVCWGG QRGGL PGNTLVWNSP FLVCWGG QRGGL PGNTLWNSSA SLVCWNG QRGGL RGBSLWNTD KNVCWEG QRGGL RGBSLWNTD KNVCWEG QRGGL RGBSLWNTD KNVCWEG QRGGL RGBSLWNTS HRVCWNG QRGGL RGBSLWNTS QRVCWGG QRGGL RGBSLWNTS QRVCWGG QRGGL RGRECLNRLG VKVCWEG QRGGL PGNTLNNTG TLVCWGG QLGGL VGNKVVHSMDN KVVCWEG QLGGL VTDDR IENSTA QRVCWQG QLGGL TDDR IENSTA QRVCWQG QLGGL PDSDLWVDNSA SIACWRG QLGGL DFJGJYLSFFG FSTQN QRGGL PDSDLWDDRS FSTQN	GLWEDQIYCLDMS	DGPTCWNG	QDGGL
MRDQIYCLDMSNGPTCWNGQDRGL      WRDQIYCLDMSRGPTCWNGQDGGL      LWRDQIYCLDMSAGPTCWNGQDGGL      LWRDQIYCLDMAAGPTCWNGQDGGL      LWRDQIYCLDMAAGPTCWNGQDGGL      LFGNTLVNNSP	GLWRDQIYCLDMT	NGPTCWNG	QDGGL
MRDQ IYCLDMS NGPTCWNG QDGGL      WKDQ IYCLDMT CAPTCWNG QDGGL      /EGNTLVNNS P YLVCWDG QKGGL      /EGNTLVNNS P	GLWRDQIYCLDMT	NGPTCWNG	QDRGL
WKDQ IYCLDMT	GLWRDQIYCLDMS	NGPTCWNG	QDGGL
WRDD[YYCLDMAAGPTCWNGQKGGL    PEGNTLVNNSPVLVCWDGQKGGL    PEGNTLVNNSPVLVCWNGQKGGL    PLEDRLESIDPNQLVCWNGKLEDGRMRQKGWS    QGDSLINNTEHRVCWEGQAGGL    RAGDSLWNTDKNVCWEGQAGGL    RAGDSLWNTDRVVCWEGQAGGL    RAGDSLWNTD	GIWKDQIYCLDMT	EGPTCWNG	QDGGL
JEGNTILVINISP	GIWRDQIYCLDMA	AGPTCWNG	QDGGL
FEGDTLTMSSA  SLVCWNGC	SVEGNTLVNNSP	YLVCWDG	QKGGL
LEDRLESIDENCLVCWNGKLEDGRMRQKGWS    QGBDSLINNTEHVVCWEGQAGGL    RRGDSLWNTD	HVEGDTLTNSSA	SLVCWNG	QKGGL
VQGB2LINRTE	IPLEDRLESIDPN	QLVCWNGKLE	DGRMRQKGWS
/RGBSLWNTTDKNVCWEGQAGGL    RRGDSLWNTTDKNVCWEGQAGGL    RRMGDLCNSTK	MVQGDSLINRTE	HMVCWEG	QKGGL
RGB5LVNTTD	DVRGDSLVNTTD	KMVCWEG	QAGGL
ERNGTLCNSTKHRVCWNGQKGGL    FRINGTLCNSTK	DVRGDSLVNTTD	KIVCWEG	QAGGL
ERNGTLCNSTK	TIRNGTLCNSTK	HRVCWNG	QKGGL
HINTILINSTSQPVCWQGQEGGL    JQGNSLVNNSNQRVCWQGQKGGL    JRGRECLNRLGVKVCWEGQKGGL    JRGRECLNRLGVKVCWEGQKGGL    JRGRECLNKLGVKVCWEGQKGGL    JRGRECLNKLGVKVCWEGQKGGL    JRGRECLNKSGVKVCWEGQKGGL    JRGRECLNKSGVKVCWEGQKGGL    JRGRECLNKSGVKVCWEGQKGGL    JQGNTLUNKTGVKVCWEGQKGGL    JQNNTLQNRGNQKVCWQGQLGGL    JQNNTLQNRGNQRVCWQGQLGGL    JDDDRIENSTAQRVCWQGQLGGL    JEDGTLLNKTKQKGGL    JEDGTLLNKTK	TIRNGTLCNSTK	HRVCWNG	QKGGL
VQGNSLVNNSNQRVCVNGGQAGGL      VRGRECLNRLGVKVCWEGQKGGL      /RGRECLNRLGVKVCWEGQKGGL      /RGRECLNRLGVKVCWEGQKGGL      /RGRECLNRLGVKVCWEGQKGGL      /RGRECLNRLGVKVCWEGQKGGL      /RGRECLNRLGVKVCWEGQKGGL      /QGRHCLNKTG	RVHNNTLINSTS	QPVCWQG	QEGGL
/RGRECLNRLG	KVQGNSLVNNSN	QRVCWNG	QAGGL
RGRECLNRLGVKVCWEG   QKGGL      /RGRECLNRLGVKVCWEG   QKGGL      /RGRECLNRLGVKVCWEG	MVRGRECLNRLG	VKVCWEG	QKGGL
//RGRECLNRLG	MVRGRECLNRLG	VKVCWEG	QKGGL
//RGRECLNRLG	MVRGRECLNRLG	VKVCWEG	QKGGL
RGRECLNKLG	MVRGRECLNRLG	VKVCWEG	QKGGL
/RGNECLNKSG	MVRGRECLNKLG	VKVCWEG	QKGGL
JQGRICLNKTG	MVRGNECLNKSG	VKVCWEG	QKGGL
JVGNEVUSMDN	MVQGRHCLNKTG	IKVCWEG	QKGGL
//RNKVWHSKGDDTVCWEGCKGL      /QNNTLQNRGNEIVCWNGQLGGL      /QNNTLQNRGN	RVVGNEVVSMDN	KKVAWEG	OKGGL
/QNNTLQNRGNEIVCWNGQLGGL      /TDDRIENSTAQRVCWQGQLGGL      /TDDRIENSTAQRVCWQGQLGGL      /EDGTLLNKTKDTVCWYGQLGGL      /EDGTLLNKTKDTVCWYGQKGGL      LIDGTLSKDA	YVRNKVVHSKGD	DIVCWEG	OKGGL
\TDDRIENSTAORVCWQGOlGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOLGGL      \TDDRIENSTAORVCWQGOKGGL      \TDDRIENSTAORVCWQGOKGGL      \TDDRIENSTAORVCWQGOKGGL      \TDDRIENSTAORVCWGGOKGGL      \TDDRIENSTAORVCWGG	MVONNTLONRGN	EIVCWNG	OLGGL
/TDDRIENSTA   ORVCWQG   OLGGL      /TDDRIENSTA   ORVCWQG   OLGGL      /EDGTLINKTK	TVTDDRIENSTA	QRVCWQG	QLGGL
/TDDRIENSTA	TVTDDRIENSTA	ORVCWOG	ÕLGGL
/EDGTLLNKTKDTVCWYGQDGGT    /HNCKLVSKDNNLVCWYGQKGGL    /HNCKLVSKDN	TVTDDRIENSTA	ORVCWOG	QLGGL
/HNGKLVSKDNNLVCWQGQKGGL IIDGIIQSTFGFLVCW3GQLGGF PDSDLQWVDDKSGKSYEGHRGF PDSDLQWVDDKSGKSYEGHRGF HKGDLVSEPGCYRGHQGM DPDYGVEFEDEGCYRGHQGM DPDTGIMVDNNVCWIDDESGK DPDTGIMVDNPWSRTGDESGK SHFTRILECKALGLDAPHTWADGVFSGL JWFFKARYEEALALGIEAPHTWADGVFSGL WGGG YD	KVEDGTLLNKTK	DTVCWYG	ÔDGGL
IIDGIIQSTFGFLVCWSGOLGGF IEGGEIVNAGGSIACWRGQAGGL TSDSLQWYDDKSGKSYEGHIRGF IHKGFLVSEPGKYEGHIRGF DPVYGVFPDGTWAWAGDESCK DPVTGVFPDGTWAWAGDESCK SHPTRILEECRLHGIEAPHVWADGVFSGL JAHFTRILDECTALGLDTPHTWADGVFSGL JVTFKARYEEALALGIEAPHVWADGVFSGL VFKARYEEALALGIEAPHVWADGAFSGL V G G Y D	EVHNGKLVSKDN	NLVCWOG	ÔKGGL
EEGGEIVNAGGSIACWRGQAGGL      DSDLQWVDDKSGKSYEGHIRGF      HKGDLUSSPFSFQNHQGGM      PDQYGEFIEDEGCYRGHGGI      DPUYGVFPDGTWANAGDESGK      PDPTGIMVDNNVCWIDDESGK      SHFTRILEECRLHCIEAPHVWADGVFSGL      AKFYERIGLECHLGIEAPHTWADGVFSGL      VTFKARYEEALALGIEAPHVWADGAFSGL      W G    G      Y D	EIIDGIIOSTFG	FLVCWSG	OLGGE
PSDLQWVDDKSGKSYEGHIGG HKGDLUSEPGCYRGHQGM PQYGYEFEDEGCYRGHQGI ADPVYGVFPDGWAWAGDESGK JDPDTGIMVDNPWSRTGDESGK JSHFTRILEECKLHGIEAPHVWADGVFSGL JAHFTRILECTALGLDTPHIWADGVFSGL JVFFKARYEEALALGIEAPHVWADGVFSGL JVFFKARYEEALALGIEAPHVWADGAFSGL Y D	KIEGGEIVNAGG	STACWRG	OAGGI
HKGDLUSSEPFSPQNHQGGM PQYGEFIEDEGVRGHGGI DPVYGVFPDGTWAWAGDESGK SDPUTGIWDDNNVCWIDDGAGK IS-NNQLTAQSPWSRTGDESGK SHFTRILEECRLHGIEAPHVWADGVFSGL UKFYDRYQKILKDLGIDAPHTWGDGVFSGL VTFKARYEEALALGIEAPHVWADGAFSGL V G Y D	KEDSDLOWVDDKS	GKSYEG	HIRGE
DQVGEFIEDEGCVRGHQGU DDPUGVEFIEDEGCVRGHLGGI DDPUTGIMVDNPWSRTGDGAGK S-NNQLTAQSPWSRTGVFSGL JAHFTRILDECTALGLDTPHIWADGVFSGL JAHFTRILDECTALGLDTPHIWADGVFSGL JVTFKARYEEALALGIEAPHVWADGAFSGL W G G Y D	DIHKGDLLVSEP	FSFON	HOGGM
DPVYGVPPDGTWAWAGDESCK DPDTGIWUDNPWSRTGDESCK SHTTRILEECRLHGIEAPHVWADGVFSGL JAHFTRILDECTALGLDTPHTWADGVFSGL JVTFKARYEEALALGIEAPHTWGDGVFSGL VTFKARYEEALALGIEAPHTWADGAFSGL W G G Y D	KEDOYGEFTEDE	GCYRG	HLGGT
NUTRICULDS IN A NUTRICDUSING DPDTGINVDNNUTRICDESGK LSHFTRILEECRLHGIEAPHVWADGVFSGL LKFFKDRYQKLKDLGIDAPHTWGDGVFSGL JVTFKARYEEALALGIEAPHVWADGAFSGL W G G Y D	SADRVVGVERDC	TWAWAG	DESCK
SINDLIAQSPWSRTGDGANK SHFTRILECRIHGIEAPHVWADGVFSGL JAHTRILDECRIGIDAPHTWADGVFSGL JXFFXRYQKLKDLGIDAPHTWGDGVFSGL JVTFKARYEEALALGIEAPHVWADGAFSGL W G G Y D	TPDPDTGTMUDN		DGAGV
SHFTRILEECRLHGIEAPHVWADGVFSGL JAHFTRILEECRLHGIEAPHVWADGVFSGL JKFFRRYQKIKLDGIDAPHTWGDGVFSGL VTFKARYEEALALGIEAPHVWADGAFSGL W G G Y D	VIS-NNOLWAOS		DGAGA
JAHTTRILDECTALGLDPPHIWADGVFSGL JAHTTRILDECTALGLDPPHIWADGVFSGL JXTFKARYEEALALGIEAPHVWADGAFSGL W G G Y D	A TO - MUCHTAGO		DEAGK
AMERIKALDSLDTPHIWALDSVFSGL KKEYEDRYQKIKLDLGIDAPHTWGDGVFSGL JVTFKARYEEALALGIEAPHVWADGAFSGL W G G Y D	DESERTRIESECREEC	JIEAPHVWADG	vr SGL
JAKFRURYEEALALGIDAPHTWODGVFSGL JVTFKARYEEALALGIEAPHVWADGAFSGL WGGG YD	NLARPTRILDECTALC	SEDTPHIWADG	VFSGL
SVIFRARIEEALALGIEAPHVWALGAFSGL W G <i>G</i> Y D	DIMERAPRICAL	STDAPHTWGDG	VFSGL
WG G YD	DEVIFKARIEEALALO	JEAPHVWADG	AFSGL
Y D		wg	G
		Y D	
F		Ϋ́D F	

Fig. 2. Alignment of the conserved block III of 56 L polymerase sequences from the Rhabdoviridae. Amino acids belonging to the same residue family (Poch et al., 1990), conserved in at least 54 of the 56 sequences and among the different viral families, are described in the text. Residues which are conserved in the Rhabdoviridae are shown in roman type. Residues conserved in all the Rhabdoviridae and the Paramyxovirinae subfamily are shown in italics. Amino acids conserved in polymerases of other unsegmented (-) RNA viruses are shown in bold type. Residues conserved in other RNA polymerases are shown in bold and underlined type (Dhillon et al., 2000; Le Mercier et al., 1997; Müller et al., 1994; Vieth et al., 2004).

to as the 'Le Dantec' group, was also seen to form a distinct cluster, although with only 66% quartet puzzling support. Finally, the phylogenetic position of Tibrogarganvirus was ambiguous. While there was weak support (54% quartet puzzling) for this virus clustering with the Le Dantec group, the length of the branch leading to Tibrogarganvirus implies that it should be classified in its own group, although this contains only a single virus at present.

Perhaps the most notable result from our phylogenetic analysis was the strong support (98%) for a virus 'supergroup', herein named 'dimarhabdovirus' (sigla for 'dipteran-mammal associated rhabdovirus'). This contained the four new groups of viruses described above, as well as the Vesiculovirus and Ephemerovirus genera. Despite major differences in genome organization, ephemeroviruses and vesiculoviruses share many similar biological characteristics. They are, together with the other dimarhabdoviruses, the only recognized rhabdovirus genera with viruses that replicate in both vertebrate and invertebrate hosts, and have biological cycles involving transmission by hematophagous dipterans. Although there is strong phylogenetic support for the dimarhabdovirus supergroup, the precise branching order within this group cannot be resolved on the L polymerase data. Indeed, there is a clear need for further phylogenetic studies within the dimarhabdovirus supergroup, particularly with respect to the demarcation of genera, which currently seems to be influenced more by genome structure than host-vector relationships. For example, compared to vesiculoviruses, ephemeroviruses contain multiple additional ORFs, including a second glycoprotein gene  $(G_{NS})$  that appears to have been acquired by gene duplication (Wang & Walker, 1993). There is some evidence that Flanders virus may also have a complex pattern of gene expression (Boyd & Whitaker-Dowling, 1988). Although the functions of these additional proteins are not understood, revealing the evolution of genome complexity may be an important factor in resolving the taxonomy of this supergroup.

In sum, the sort of molecular phylogenetic analysis undertaken here, especially if combined with data on genome organization, is likely to provide a more useful guide to taxonomic classification, particularly for assignments above the species level and even among all (-) RNA viruses (Vieth et al., 2004). Indeed, our phylogenetic analysis of a conserved L-gene segment appears to provide a useful taxonomic tool for the rapid classification of rhabdoviruses.

#### Association between phylogenetic relationships and mode of transmission

A number of important biological conclusions can be drawn from the rhabdovirus phylogeny presented here. First, assuming a mid-point rooting of the tree, there is major split between those viruses that infect fish (novirhabdoviruses) and plants and which employ arthropods as vectors (cytorhabdoviruses and nucleorhabdoviruses), and those viruses that mainly infect mammals, lizards and dipterans (dimarhabdoviruses). Such a division illuminates the biology of a number of key rhabdoviruses. For example, although vesicular stomatitis virus (VSV) is responsible for a disease of horses, cattle and pigs and can be transmitted directly by transcutaneous or transmucosal routes (Stallknecht et al., 1999), there is good evidence that VSV may be an insect virus (Rodriguez, 2002). Indeed, it has been found to replicate in biting midges (Culicoides) and Simulium blackflies (Mead et al., 1999), and has been



**Fig. 3.** Phylogenetic relationships of the *Rhabdoviridae* based on a maximum-likelihood analysis of a 158-residue alignment of the L polymerase region. The established rhabdovirus genera as well as the new groups proposed here are indicated. Horizontal branches are drawn to scale and quartet puzzling frequencies are shown for key nodes (values in italics are for genera, groups and supergroups, while all other quartet puzzling frequencies are shown in roman type). The tree is mid-point rooted for purposes of clarity only, and all potential outgroup sequences were deemed too divergent to include in the analysis.

isolated from sandflies, and epidemic and endemic bursts depend on region, season and the presence of dipterans (*Lutzomya, Simulidae, Culicoides* and *Musca domestica*) (Gard *et al.*, 1984; Walker & Cybinski, 1989). All these factors suggest that VSV may be insect-borne. Similarly, *Bovine ephemeral fever virus*, which is frequently found in Australasia, Asia and Africa, is also dipteran-transmitted, using vectors such as biting midges and culicine and anopheline mosquitos. Finally, viruses assigned by our phylogenetic analysis to the four new groups (the Le Dantec, Tibrogargan, Hart Park and Almpiwar groups) were all found to infect dipterans and in some cases mammals (Tibrogargan, Le Dantec and Ngaingan viruses) and lizards (Charleville virus) also.

Importantly, there is as yet no evidence for a virus that would constitute a link between plant and fish viruses and dimarhabdoviruses and the lyssaviruses. Furthermore, the uncertainty over branching order at the root of the tree makes it difficult to determine whether the ancestral mode of transmission in the rhabdoviruses was vector or nonvector transmission. A similar lack of resolution at the base of tree was found in a previous phylogenetic analysis of six genera of rhabdoviruses (Vieth *et al.*, 2004). However, the major phylogenetic division between these groups indicates that the biology of the rhabdoviruses could be strongly influenced by mode of transmission and by the host (plant, fish or mammal) and vector (orthopteran, homopteran or dipteran) species. Similar findings have been reported in other RNA viruses, such as the flaviviruses (Gaunt *et al.*, 2001) and the tick-borne nairoviruses (Honig *et al.*, 2004).

Finally, it is noteworthy that levels of genetic diversity vary substantially among genera. This is most apparent when comparing the tightly clustered lyssaviruses (the different genotypes of which our phylogenetic analysis cannot easily distinguish) with the cytorhabdoviruses and nucleorhabdoviruses, which are highly diverse. Indeed, the entire *Lyssavirus* genus, although clearly separate from the other rhabdoviruses, is less divergent than two serotypes (Indiana and New Jersey) of VSV. The most likely explanation for such differences is that these genera differ substantially in age, with the lyssaviruses evolving most recently. However, it is also possible that strong selective constraints acting against sequence change in the lyssaviruses also serve to limit amino acid variation (Guyatt *et al.*, 2003; Holmes *et al.*, 2002; Kissi *et al.*, 1999).

# ACKNOWLEDGEMENTS

The authors would like to thank Hervé Zeller for helpful discussions, and Jocelyn Thonnon for having provided some of the isolates stored at the collection of arboviruses, and The Wellcome Trust for financial support. The technical assistance of Christine Dimmock is gratefully acknowledged. We would like to dedicate this paper to the memory of Dr Jean Pierre Digoutte, former Director of the Institut Pasteur of Dakar, Senegal.

### REFERENCES

Arai, Y. T., Kuzmin, I. V., Kameoka, Y. & Botvinkin, A. D. (2003). New lyssavirus genotype from the lesser mouse-eared bat (*Myotis blythi*), Kyrghyzstan. *Emerg Infect Dis* 9, 333–337.

Badrane, H. & Tordo, N. (2001). Host switching in *Lyssavirus* history from the Chiroptera to the Carnivora orders. J Virol 75, 8096–8104.

**Barr, J., Chambers, P., Pringle, C. R. & Easton, A. J. (1991).** Sequence of the major nucleocapsid protein gene of pneumonia virus of mice: sequence comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses. *J Gen Virol* **72**, 677–685.

**Basurco, B. & Benmansour, A. (1995).** Distant strains of the fish rhabdovirus VHSV maintain a sixth functional cistron which codes for a nonstructural protein of unknown function. *Virology* **212**, 741–745.

Basurco, B., Vende, P., Monnier, A. F., Winton, J. R., de Kinkelin, P. & Benmansour, A. (1995). Genetic diversity and phylogenetic classification of viral hemorrhagic septicemia virus (VHSV). *Vet Res* 26, 460–463.

Bock, J. O., Lundsgaard, T., Pedersen, P. A. & Christensen, L. S. (2004). Identification and partial characterization of Taastrup virus: a newly identified member species of the Mononegavirales. *Virology* **319**, 49–59.

Bourhy, H. & Sureau, P. (1991). Laboratory Methods for Rabies Diagnosis. Paris: CLRE, Institut Pasteur.

Bourhy, H., Kissi, B. & Tordo, N. (1993). Molecular diversity of the Lyssavirus genus. *Virology* 194, 70–81.

Boyd, K. R. (1972). Serological comparisons among Hart Park virus and strains of Flanders virus. *Infect Immun* 6, 933–937.

Boyd, K. R. & Whitaker-Dowling, P. (1988). Flanders virus replication and protein synthesis. *Virology* 163, 349–358.

Calisher, C. H., Karabatsos, N., Zeller, H., Digoutte, J. P., Tesh, R. B., Shope, R. E., Travassos da Rosa, A. P. & St George, T. D. (1989). Antigenic relationships among rhabdoviruses from vertebrates and hematophagous arthropods. *Intervirology* **30**, 241–257.

**Crysler, J. G., Lee, P., Reinders, M. & Prevec, L. (1990).** The sequence of the nucleocapsid protein (N) gene of Piry virus: possible domains in the N protein of vesiculoviruses. *J Gen Virol* **71**, 2191–2194.

Dale, J. L. & Peters, D. (1981). Protein composition of the virions of five rhabdovirus. *Intervirology* 16, 86–94.

Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng* 3, 461–467.

**Dhillon, J., Cowley, J. A., Wang, Y. & Walker, P. J. (2000).** RNA polymerase (L) gene and genome terminal sequences of ephemeroviruses bovine ephemeral fever virus and Adelaide River virus indicate a close relationship to vesiculoviruses. *Virus Res* **70**, 87–95.

Elliott, R. M., Dunn, E., Simons, J. F. & Pettersson, R. F. (1992). Nucleotide sequence and coding strategy of the Uukuniemi virus L RNA segment. *J Gen Virol* **73**, 1745–1752.

Gard, G. P., Cybinski, D. H. & Zakrzewski, H. (1984). The isolation of a fourth bovine ephemeral fever group virus. *Aust Vet J* 61, 332.

Gaunt, M. W., Sall, A. A., de Lamballerie, X., Falconar, A. K. I., Dzhivanian, T. I. & Gould, E. A. (2001). Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol* 82, 1867–1876.

Guyatt, K. J., Twin, J., Davis, P., Holmes, E. C., Smith, G. A., Smith, I. L., Mackenzie, J. S. & Young, P. L. (2003). A molecular epidemiological study of Australian bat lyssavirus. *J Gen Virol* **84**, 485–496.

Heaton, L. A., Hillman, B. I., Hunter, B. G., Zuidema, D. & Jackson, A. O. (1989). Physical map of the genome of Sonchus yellow net virus, a plant rhabdovirus with six genes and conserved junction sequences. *Proc Natl Acad Sci U S A* 86, 8665–8668.

Holmes, E. C., Woelk, C. H., Kassis, R. & Bourhy, H. (2002). Genetic constraints and the adaptive evolution of rabies virus. *Virology* 292, 247–257.

Honig, J. E., Osborne, J. C. & Nichol, S. T. (2004). The high genetic variation of viruses of the genus *Nairovirus* reflects the diversity of their predominant tick hosts. *Virology* **318**, 10–16.

Jin, H. & Elliott, R. M. (1991). Expression of functional Bunyamwera virus L protein by recombinant vaccinia viruses. *J Virol* 65, 4182–4189.

Jin, H. & Elliott, R. M. (1992). Genesis of the L protein encoded by Bunyamwera virus and production of monospecific antibodies. *J Gen Virol* 73, 2235–2244.

Kemp, G. E., Lee, V. H., Moore, D. L., Shope, R. E., Causey, O. R. & Murphy, F. A. (1973). Kotonkan, a new rhabdovirus related to Mokola virus of the rabies serogroup. *Am J Epidemiol* 98, 43–49.

Kissi, B., Tordo, N. & Bourhy, H. (1995). Genetic polymorphism in the rabies virus nucleoprotein gene. *Virology* 209, 526–537.

Kissi, B., Badrane, H., Audry, L., Lavenu, A., Tordo, N., Brahimi, M. & Bourhy, H. (1999). Dynamics of rabies virus quasispecies during serial passages in heterologous hosts. *J Gen Virol* 80, 2041–2050.

Kuzmin, I. V., Orciari, L. A., Arai, Y. T., Smith, J. S., Hanlon, C. A., Kameoka, Y. & Rupprecht, C. E. (2003). Bat lyssaviruses (Aravan and Khujand) from Central Asia: phylogenetic relationships according to N, P and G gene sequences. *Virus Res* **97**, 65–79.

Landes-Devauchelle, C., Bras, F., Dezelee, S. & Teninges, D. (1995). Gene 2 of the sigma rhabdovirus genome encodes the P protein, and gene 3 encodes a protein related to the reverse transcriptase of retroelements. *Virology* 213, 300–312.

Le Mercier, P., Jacob, Y. & Tordo, N. (1997). The complete Mokola virus genome sequence: structure of the RNA-dependent RNA polymerase. *J Gen Virol* 78, 1571–1576.

Masters, P. S. & Banerjee, A. K. (1987). Sequences of Chandipura virus N and NS genes: evidence for high mutability of the NS gene within vesiculoviruses. *Virology* 157, 298–306.

McWilliam, S. M., Kongsuwan, K., Cowley, J. A., Byrne, K. A. & Walker, P. J. (1997). Genome organization and transcription strategy in the complex GNS-L intergenic region of bovine ephemeral fever rhabdovirus. *J Gen Virol* **78**, 1309–1317.

Mead, D. G., Maré, C. J. & Ramberg, F. B. (1999). Bite transmission of vesicular stomatitis virus (New Jersey serotype) to laboratory mice by *Simulium vittatum* (Diptera: *Simulidae*). *Entomol Soc Am* 36, 410–413.

Müller, R., Poch, O., Delarue, M., Bishop, D. H. & Bouloy, M. (1994). Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J Gen Virol* **75**, 1345–1352.

Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* 8, 3867–3874.

Poch, O., Blumberg, B. M., Bougueleret, L. & Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* 71, 1153–1162.

Rodriguez, L. L. (2002). Emergence and re-emergence of vesicular stomatitis in the United States. *Virus Res* 85, 211–219.

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.

Schnell, M. J. & Conzelmann, K. K. (1995). Polymerase activity of *in vitro* mutated rabies virus L protein. *Virology* 214, 522–530.

**Shope, R. (1995).** Rabies virus antigenic relationships. In *The Natural History of Rabies*, vol. 1, pp. 141–152. Edited by G. Baer. Berlin: Springer.

**Sleat, D. E. & Banerjee, A. K. (1993).** Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase. *J Virol* **67**, 1334–1339.

Stallknecht, D. E., Howerth, E. W., Reeves, C. L. & Seal, B. S. (1999). Potential for contact and mechanical vector transmission of vesicular stomatitis virus New Jersey in pigs. *Am J Vet Res* **60**, 43–48.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673–4680.

Tomori, O., Fagbami, A. & Kemp, G. K. (1974). Kotonkan virus: experiment infection of white Fulani calves. *Bull Epizoot Dis Afr* 22, 195–200.

Tordo, N., Poch, O., Ermine, A., Keith, G. & Rougeon, F. (1988). Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses. *Virology* 165, 565–576.

Tordo, N., Benmansour, A., Calisher, C. & 7 other authors (2004). Rhabdoviridae. In *Virus Taxonomy, VIIIth Report of the ICTV*, pp. 623–644. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier/Academic Press. Vieth, S., Torda, A. E., Asper, M., Schmitz, H. & Günther, S. (2004). Sequence analysis of L RNA of Lassa virus. *Virology* 318, 153–168.

Walker, P. J. & Cybinski, D. H. (1989). Bovine ephemeral fever and rhabdoviruses endemic to Australia. *Aust Vet J* 66, 398–400.

Walker, P. J., Byrne, K. A., Riding, G. A., Cowley, J. A., Wang, Y. & McWilliam, S. (1992). The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. *Virology* 191, 49–61.

Walker, P. J., Wang, Y., Cowley, J. A., McWilliam, S. M. & Prehaud, C. (1994). Structural and antigenic analysis of the nucleoprotein of bovine ephemeral fever rhabdovirus. *J Gen Virol* 75, 1889–1899.

Wang, Y. & Walker, P. J. (1993). Adelaide River rhabdovirus expresses consecutive glycoprotein genes as polycistronic mRNAs: new evidence of gene duplication as an evolutionary process. *Virology* 195, 719–731.

Wang, Y., McWilliam, S. M., Cowley, J. A. & Walker, P. J. (1994). Complex genome organization on the  $G_{NS}$ -L intergenic region of Adelaide River rhabdovirus. *Virology* **203**, 63–72.

Wang, Y., Cowley, J. A. & Walker, P. J. (1995). Adelaide River virus nucleoprotein gene: analysis of phylogenetic relationships of ephemeroviruses and other rhabdoviruses. *J Gen Virol* 76, 995–999.

Wetzel, T., Dietzgen, R. G. & Dale, J. L. (1994). Genomic organization of lettuce necrotic yellows rhabdovirus. *Virology* 200, 401–412.

Xiong, Y. & Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 10, 3353–3362.