Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 111 (2005) 28–43
Virus Research
www.elsevier.com/locate/virusres

Phylogenetic relationships of Irkut and West Caucasian bat viruses within the *Lyssavirus* genus and suggested quantitative criteria based on the N gene sequence for lyssavirus genotype definition

Ivan V. Kuzmin [a,b,*], Gareth J. Hughes [a], Alexandr D. Botvinkin [c], Lillian A. Orciari [a], Charles E. Rupprecht [a]

[a] *Centers for Disease Control and Prevention, 1600 Cliffton Road, Atlanta, GA 30333, USA*
[b] *Institute for Natural Foci Infections, 7 Prospekt Mira, Omsk 644080, Russia*
[c] *Plague Control Research Institute of Siberia and Far East, 78 Trilissera Street, Irkutsk 664047, Russia*

Available online 8 April 2005

Abstract

The nucleoprotein (N), phosphoprotein (P) and glycoprotein (G) genes of Irkut and West Caucasian bat viruses (WCBV) were sequenced and compared with those of other lyssaviruses. N gene nucleotide identities provided unequivocal separation of all lyssavirus genotypes with an identity threshold of 82%. On this basis, Irkut virus should be considered as a new genotype with particular relatedness to genotypes 4 and 5 (78.0–78.6% identity for N gene nucleotides and 90.4–92.6% for amino acids). Furthermore, genotypes 4–6, together with Aravan, Khujand and Irkut viruses, present a solid phylogroup of Old World bat lyssaviruses. This relationship is apparent using all three viral genes, and causes overlap between intragenotype and intergenotype identities for the P gene (Aravan, Khujand viruses and genotype 6) and for the G gene (Aravan, Khujand, genotypes 5 and 6). WCBV is the most divergent of known lyssaviruses with only limited relatedness to genotypes 2 and 3.

Keywords: *Rhabdovirus*; *Lyssavirus*; Rabies; *Chiroptera*; Bat; Phylogeny; Genotype

1. Introduction

Until recently, the *Lyssavirus* genus, family *Rhabdoviridae*, was believed to consist of seven distinct genotypes (GTs). *Rabies virus* (RABV; GT 1), is distributed worldwide among terrestrial mammals and bats, presents the most comprehensive collection of isolates, and has been extensively studied, due to its health and economic significance. Lagos bat virus (LBV; GT 2) was first isolated in Nigeria from the frugivorous bat (*Eidolon helvum*) in 1956 (Boulger and Porterfield, 1958). This virus was also isolated from the bat *Micropteropus pussilus* in the Central Africa Republic (Sureau et al., 1980), from the bat *Epomophorus wahlbergi* in South Africa (Meredith, 1980), cats in South Africa (King and Crick, 1988) and Zimbabwe (Foggin, 1988), and from a dog in Ethiopia (Mebatsion et al., 1992). Mokola virus (MOKV; GT 3) was first isolated from shrews in Nigeria in 1968 (Shope et al., 1970). Thereafter, MOKV was detected in shrews from Nigeria and Cameroon, humans in Nigeria, domestic cats in Zimbabwe, Ethiopia and South Africa, a domestic dog in Zimbabwe and a rodent (*Lophuromys sikapusi*) from the Central Africa Republic (reviewed by Nel et al., 2000). Duvenhage virus (DUVV, GT 4) was isolated from a human, who died after a bat bite in 1970 in South Africa (Meredith et al., 1971). It has also been identified in the insectivorous bat (*Miniopterus sp.*) in South Africa and another bat species (*Nycteris thebaica*) in Zimbabwe (Van der Merwe, 1982; King and Crick, 1988). Since its discovery in 1968, European bat lyssavirus, type 1 (EBLV-1; GT 5) has been isolated from a number of European countries, where it was suspected since the 1950s (Schneider et al., 1985). The primary host species for the virus is considered the serotine bat (*Eptesicus serotinus*) (Amengual et al., 1997;

* Corresponding author. Fax: +1 404 639 1564.
E-mail address: ibk3@cdc.gov (I.V. Kuzmin).

0168-1702/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2005.03.008

Serra-Cobo et al., 2002). One human case of EBLV-1 infection occurred following a bat bite in Russia in 1985 (Selimov et al., 1989). European bat lyssavirus, type 2 (EBLV-2; GT 6) was isolated in Finland from a biologist, who died of rabies (Lumio et al., 1986). Later, it was isolated from bats (primarily from the genus *Myotis*) in northwestern Europe (Amengual et al., 1997), and more recently caused a second recorded human death, this time in the United Kingdom (Fooks et al., 2002). Phylogenetic analysis has demonstrated that EBLV-1 and EBLV-2 form relatively homogenous clusters, but each may be additionally subdivided into two lineages, "a" and "b" (Amengual et al., 1997). Australian bat lyssavirus (ABLV; GT 7) was discovered in 1996. Initially it was isolated from pteropid bats, and two humans that died of rabies following bat exposure. Later ABLV was identified from five different bat species with at least two separate lineages originating from frugivorous and insectivorous species (Fraser et al., 1996; Gould et al., 2002; Guyatt et al., 2003).

Clearly, most non-rabies lyssaviruses are associated with bats. The ecology of MOKV has not been sufficiently studied, so it is not possible to determine the principal host or suggest inferences regarding its epidemiology. Shortly after the discovery of LBV and MOKV, Africa was considered as the probable birth place of the *Lyssavirus* genus, and the *Chiroptera* were suggested as the first affected order on the route of adaptation from plant and arthropod rhabdoviruses to mammalian hosts (Shope, 1982). Furthermore, since EBLV-1 had been demonstrated to be related to DUVV, an introduction of the virus from Africa to Europe with migrating bats was hypothesised (Schneider et al., 1985; Amengual et al., 1997; Serra-Cobo et al., 2002).

Very limited information on bat lyssaviruses is available for Asia. Only a few records reported presumable rabies virus isolates of Chiropteran origin in India (Pal et al., 1980) and Thailand (Smith et al., 1968), but these have never been confirmed by further identification or other observations. The presence of antibodies to ABLV was demonstrated in sera from bats in the Philippines, however, no isolates were obtained (Arguin et al., 2002).

Two lyssaviruses have been isolated from bats in Central Asia. Aravan virus was isolated in southern Kyrgyzstan in 1991 (Kuzmin et al., 1992). Nucleotide sequencing of the entire nucleoprotein (N) gene has provided an indication that Aravan virus might be regarded as a new lyssavirus genotype with distant relatedness to GT 4 and GT 5 (Arai et al., 2003). A second example, Khujand virus, was isolated in Northern Tajikistan during 2001 (Kuzmin et al., 2002). Further phylogenetic analysis of the N, phosphoprotein (P) and glycoprotein (G) genes of these viruses suggests that Khujand virus is most related to genotype 6, while Aravan virus is related to Khujand virus, and demonstrated moderate similarity to GTs 4–6 (Kuzmin et al., 2003).

During 2002, two new lyssaviruses have been isolated from Eurasian bats. According to preliminary identification with anti-nucleocapsid monoclonal antibodies and comparison of limited N gene sequences, both may be considered as new genotypes of the *Lyssavirus* genus (Botvinkin et al., 2003). These viruses have been named West Caucasian bat virus (WCBV) and Irkut virus.

In this study we describe the genetic properties of Irkut virus and WCBV, determine their phylogenetic relationships according to the entire N, P and G gene sequences, and discuss quantitative approaches for lyssavirus genotype definition considering the bat lyssaviruses discovered during recent years.

2. Materials and methods

2.1. Virus isolates

The isolation of Irkut virus and WCBV has been described earlier (Botvinkin et al., 2003). Briefly, Irkut virus was isolated in the town of Irkutsk (East Siberia, Baikal lake region) from the brain of a greater tube-nosed bat (*Murina leucogaster*) captured in September, 2002. WCBV was isolated from the brain of a common bent-winged bat (*Miniopterus schreibersi*) in Krasnodar region in July 2002. Both viruses were isolated by intracerebral mouse inoculation (Koprowski, 1996). In mice, both viruses caused fatal encephalitis, and typical cytoplasmic inclusions were detected in mouse brain impressions using either polyclonal or monoclonal anti-rabies virus antibodies.

For this study both viruses were passaged twice by intracerebral inoculation of mice, prior to genetic characterization.

2.2. RNA extraction, RT-PCR and nucleotide sequencing

Total RNA was extracted from infected mouse brains, using TRIzol™ (Life Technologies) according to the manufacturer's instructions. RT-PCR and direct sequencing were performed as described earlier (Sacramento et al., 1991) with modifications (Kuzmin et al., 2003). At least two RNA samples were processed with each primer pair, and both DNA strands of PCR product were sequenced at least twice.

2.3. Phylogenetic analysis

Sequence editing, alignment, consensus sequence generation and translation to amino acid sequences were performed using BioEdit software (Hall, 1999). Multiple alignment was performed using the Clustal X package (Jeanmougin et al., 1998). Neighbor joining (NJ) analysis was performed using MEGA version 2.1 (Kumar et al., 2001). Maximum parsimony (MP) analysis and maximum likelihood (ML) analysis were performed using PHYLIP version 3.6-alpha (Felsenstein, 1993). Nucleotide substitution models used transition/transversion ratios varying from 2 to 4, with empirical base frequencies, gamma distribution of rate variation among sites, and a Hidden Markov Model for inferring different rates of evolution at different sites. Bootstrap values were determined using 1000 replicates for the NJ and MP

Table 1
The lyssavirus sequences used in the present study

Year	Country	Species isolated from	Virus number	Reference	Genbank accession no.		
					N gene	P gene	G gene
GT1							
1989	USA	Raccoon	PA R89	Nadin-Davis et al. (1996)	U27221		
1992	Canada	*Lasionycteris noctivagans*	ML5	Nadin-Davis et al. (2001)	AF351834		
1988	Russia	Arctic fox	9141RUS	Kissi et al. (1995)	U22656		
1986	Burkina Fasso	Dog	88636HAV	Kissi et al. (1995)	U22486		
1990	French Guyana	Dog	9001GUY	Kissi et al. (1995)	U22478		
?	South Africa	*Viverridae spp.*	?	Jacobs et al. (unpublished)	AF467949	AF467950	
1986	Iran	Dog	8681IRA	Kissi et al. (1995)	U22482		
1982	Yugoslavia	Wolf	8653YOU	Bourhy et al. (1999)	U22482		
1991	Estonia	Raccoon dog	9339EST	Bourhy et al. (1999)	U42707		
Laboratory strain			SAD B19	Conzelmann et al. (1990)	M31046	M31046	M31046
Laboratory strain			CVS	Tordo et al. (1986)	D42112		
The same				Larson and Wunner (1990)		X55727	
The same				Morimoto et al. (1998)			AF042823
1987	USA	Raccoon	FL.RAC	Nadin-Davis et al. (2002)		AF369294	
1980	Canada	*Lasionycteris noctivagans*	4398.SHB	Nadin-Davis et al. (2002)		AF369345	
1986	Sri-Lanka	Dog	V118.DG	Nadin-Davis et al. (2002)		AF369321	
1993	Canada	Arctic fox	ARC5.RFX	Nadin-Davis et al. (2002)		AF369270	
1989	Nepal	Dog	V124.DG	Nadin-Davis et al. (2002)		AF369319	
2000	Israel	Fox	V660.FX	Nadin-Davis et al. (2002)		AF369280	
?	USA	*Lasionycteris noctivagans*	SHBRV	Morimoto et al. (1996)			U52946
?	Canada	Arctic fox	91RABN1035	Nadin-Davis et al. (1999)			U11736
?	Canada	Skunk	92RBGL0867	Nadin-Davis et al. (1999)			AF344307
1983	China	Dog/mouse brain/BHK cell culture	CTN	Tang et al. (2000)			AY009100
?	USA	Raccoon	NY516	Nadin-Davis et al. (1999)			U27214
?	Algeria	Dog	?	Benmansour et al. (1992)			M81058
?	Thailand	Dog	?	Minamoto N. (unpublished)			AB052666
GT2							
1956	Nigeria	*Eidolon helvum*	8619NGA	Kissi et al. (1995)	U22842		
1990–1991	Ethiopia	Dog	Ethlag	This paper			
1956	Nigeria	*Eidolon helvum*	v006	Nadin-Davis et al. (2002)		AF049114	
?	South Africa	*Epomophorus wahlbergii*	V267	Nadin-Davis et al. (2002)		AF049119	
1974	Central Africa Republic	*Micropteropus pussilus*	LagCAR	Badrane et al. (2001)			AF298149
1956	Nigeria	*Eidolon helvum*	LagNGA	Badrane et al. (2001)			AF298148
GT3							
1981	Zimbabwe	Cat	?	Kissi et al. (1995)	U22843		
1982	Zimbabwe	Cat	?	Le Mercier et al. (1997)	Y09762		Y09762
1990–1991	Ethiopia	Cat	Ethmok	This paper			
1998	South Africa	Cat	V635	Nadin-Davis et al. (2002)		AF369378	
1997	South Africa	Cat	V552	Nadin-Davis et al. (2002)		AF369376	
1988	Nigeria	*Crocidura sp.*	V020	Nadin-Davis et al. (2002)		AF049116	
1993	Zimbabwe	Cat	V628	Nadin-Davis et al. (2002)		AF369377	
1995	South Africa	Cat	V547	Nadin-Davis et al. (2002)		AF369374	
1970	South Africa	Cat	V241	Nadin-Davis et al. (2002)		AF049118	
1997	South Africa	Cat	V550	Nadin-Davis et al. (2002)		AF369375	
GT4							
1970	South Africa	Human	86132AS	Amengual et al. (1997)	U22848		
1970	South Africa	Human	V008	Nadin-Davis et al. (2002)		AF049115	
1981	South africa	*Miniopterus sp.*	V268	Nadin-Davis et al. (2002)		AF049120	
1970	South Africa	Human	DuvSAF1	Badrane et al. (2001)			AF298146
1981	South Africa	*Miniopterus* sp.	DuvSAF2	Badrane et al. (2001)			AF298147
GT5							
1989	France	*Eptesicus serotinus*	8918FRA	Kissi et al. (1995); Badrane et al. (2001)	U22845		AF298143

Table 1 (*Continued*)

Year	Country	Species isolated from	Virus number	Reference	Genbank accession no.		
					N gene	P gene	G gene
1985	Poland	*Eptesicus serotinus*	8615POL	Kissi et al. (1995); Badrane et al. (2001)	U22844		AF298142
1986	Denmark	*Eptesicus serotinus*	V002	Nadin-Davis et al. (2002)		AF049113	
1988	Holland	*Eptesicus serotinus*	V023	Nadin-Davis et al. (2002)		AF049117	
GT6							
1986	Finland	Human	9007FIN	Kissi et al. (1995); Badrane et al. (2001)	U22846		AF298144
1986	Holland	*Myotis dasycneme*	9018HOL	Kissi et al. (1995); Badrane et al. (2001)	U22847		AF298145
1992	Switzerland	*Myotis daubentonii*	V286	Nadin-Davis et al. (2002)		AF049121	
GT7							
1996	Australia	*Saccolaimus flaviventris*	AF081020	Gould et al. (2002)	AF081020		AF081020
1996	Australia	*Pteropus sp.*	AF006497	Gould et al. (1998)	AF006497		AF006497
1996	Australia	Human	V481	Nadin-Davis et al. (2002)		AF369373	
1996	Australia	*Pteropus sp.*	V478	Nadin-Davis et al. (2002)		AF369371	
1996	Australia	*Pteropus sp.*	V474	Nadin-Davis et al. (2002)		AF369369	
1999	Australia	*Pteropus alecto*	735-34	Guyatt et al. (2003)			AF426308
2000	Australia	*Pteropus alecto*	632-37	Guyatt et al. (2003)			AF426311
Unclassified							
1991	Kyrgizstan	*Myotis blythi*	Aravan	Kuzmin et al. (2003)	AY262023	AY262023	AY262023
2001	Tajikistan	*Myotis daubentoni*	Khujand	Kuzmin et al. (2003)	AY262024	AY262024	AY262024
2002	Russia (Irkutsk province)	*Murina leucogaster*	Irkut	This paper	AY333112	AY333112	AY333112
2002	Russia (Krasnodar region)	*Miniopterus schreibersi*	West Caucasian bat virus	This paper	AY333113	AY333113	AY333113

methods and by using 100 replicates for the ML method. The NJPLOT program from Clustal X and the TreeExplorer module of MEGA were used to obtain the graphic output. Vesicular stomatitis virus (Indiana 1 serotype; GenBank accession no. AF473864) was used as the outgroup. The other sequences used for analysis are presented in Table 1.

2.4. Positive selection analysis

Various models of codon substitution (Yang et al., 2000) implemented in the CODEML program of the PAML package (Yang, 1997), were applied with varying constraints on the values of synonymous (d_S) and nonsynonymous (d_N) substitution rates, and their ratio (ω). Models allowing for positive selection (that is, $\omega > 1$) are nested with models that do not allow for positive selection. This allows the significance of the fit for positive selection models to be tested using the likelihood ratio test. Positive selection is inferred if the positive selection model has a significantly higher likelihood than the null model and a value of $\omega > 1$ is estimated. If evidence of positive selection is suggested, Bayesian methods are used to identify which individual codons fall into the $\omega > 1$ class. Sequence alignments were trimmed to include only non-stop codons. Identical sequences were removed. For each alignment, a ML tree was generated using PAUP* 4.0, beta 10 (Swofford, 2000) by a heuristic search incorporating tree bisection-recombination (TBR) branch swapping. This tree was then used for positive selection analysis. In each case, the model of nucleotide substitution was selected using Modeltest (Posada and Crandall, 1998). Model testing using CODEML was performed as the methods of Woelk et al. (2002).

3. Results

3.1. N gene

The coding region of the entire N gene consisted of 1353 nucleotides (nt) (451 deduced amino acids) for Irkut virus and 1350 nt (450 deduced amino acids) for the WCBV. Putative phosphorylation site mapping to serine at position 389 (Dietzschold et al., 1987) was conserved. Irkut virus shared the sequence of the NIII epitope (aa 313–337) with GTs 4 and 5 (-QSC- at positions 331–333), while the WCBV shared this epitope with GT 2 and 3 (-QTC- in the same positions). With other described T- and B-cell epitopes (Ertl et al., 1991; Fu et al., 1994; da Cruz et al., 2002), Irkut virus demonstrated particular similarity to GTs 4 and 5 and Aravan virus, while the WCBV demonstrated significant diversity compared to all other lyssavirus representatives, except site 3D (aa 22–35), which was conserved amongst all sequences. For both viruses the N gene terminated with the sequence 5′-GA_7-3′ followed by -CT- for Irkut and by -CTAC- for WCBV

Table 2
Identity of the N gene and deduced nucleoprotein sequences of lyssaviruses presented in Fig. 2

Genotypes	GT 1	GT 2	GT 3	GT 4	GT 5	GT 6	GT 7	Aravan	Khujand	Irkut	WCBV
GT 1	93.7–99.9 / 82.3–99.4 [a]	81.1–85.1	76.4–82.4	87.3–88.4	87.1–88.4	86.2–87.8	89.1–92.5	88.9–90.1	89.6–90.8	85.7–87.5	80.5–81.2
GT 2	72.5–73.9	94.4 / 82.8	87.1–90.00	85.7–87.3	83.1–86.0	79.1–81.5	81.3–84.8	84.6–87.7	81.3–84.6	84.2–85.7	82.4–83.1
GT 3	69.4–73.8	76.1–78.5	94.0–98.4 / 86.1–98.7	80.6–84.6	78.8–83.3	76.0–79.7	78.8–84.0	80.8–85.3	78.4–82.2	80.4–84.6	78.4–82.4
GT 4	70.1–73.4	73.4–75.2	71.5–73.6		92.7–93.3	85.8–86.2	88.9–90.0	91.6	89.3	90.4	84.6
GT 5	70.4–74.9	74.0–74.5	69.8–73.0	78.3–79.8	98.6 / 96.5	86.4–88.0	88.2–89.8	91.6–92.0	89.1–90.2	91.7–92.6	81.7–82.4
GT 6	70.9–74.1	71.9–73.5	69.0–72.2	72.4–73.6	74.5–76.9	97.8 / 95.9	86.4–87.8	88.2–88.9	90.2–90.4	86.2	80.2
GT 7	72.4–78.0	72.5–75.1	69.7–73.1	74.2–75.7	72.5–75.6	72.1–76.0	95.6 / 84.1	91.1–92.0	90.4–92.2	87.7	81.3–82.6
Aravan	73.2–76.2	74.2–75.1	72.3–74.7	78.1	77.8–78.0	76.6–77.1	76.5		92.7	90.6	83.5
Khujand	75.3–76.5	71.9–73.6	68.2–70.9	76.0	77.0–77.3	79.0	75.8–78.5	78.8		88.0	81.1
Irkut	73.1–75.2	74.7–75.1	71.1–74.2	78.0	78.2–78.6	76.6–76.9	75.2–75.8	76.2	76.1		83.1
WCBV	72.3–73.0	74.1–74.3	70.8–72.9	73.9	72.1–72.3	72.1	73.0	73.1	70.5	72.1	

[a] Upper triangle and above the dash – amino acid identity (shaded); lower triangle and below the dash – nucleotide identity. When only one sequence of a certain genotype was available for comparison, there is no value.

as was earlier described for other lyssaviruses (Tordo et al., 1986).

Nucleotide and amino acid identities for the N gene are shown in Table 2. The maximum intrinsic diversity for the complete N gene was found in GT 1: 82.3% and 93.7% identity for the nucleotide and amino acid sequences, respectively. These data match earlier estimations based on N gene sequences (Bourhy et al., 1993). The maximum amount of homology between different GTs was shared for GTs 4 and 5 (up to 79.8 and 93.3% for nucleotide and amino acid sequences, respectively). Irkut virus was most related to GT 5 (78.6% of nucleotide and 92.6% of amino acid identity). WCBV remained distant from all other lyssaviruses however, most divergent representative of the genus according to N gene, was MOKV.

The distribution of frequencies of pairwise identities between the entire N gene nucleotide sequences (Fig. 1a) clearly distinguish intragenotype identity (82.3% and more) from intergenotype identity (79.8% and less). The amino acid sequences of the N gene provide limited resolution for genotyping (Fig. 1b), but there remains no overlap between virus groups that were defined according to nucleotide sequence identities.

A comparison of WCBV sequence identity with that of the two lyssavirus phylogroups described earlier (Badrane et al., 2001) (Table 3) show that this virus cannot be considered as a member of either of these phylogroups.

Regardless of method, results of the phylogenetic analysis demonstrated the same tree topology for all main clusters, except Aravan virus (ML and MP trees not shown). The position of Aravan virus was unstable due to its equally moderate homology with Khujand virus and with the clade joining GTs 4 and 5 (Kuzmin et al., 2003). The same topology was obtained for the 450 deduced amino acids (Fig. 2). Irkut virus was positioned as a member of the cluster joining GTs 4 and 5 (bootstrap support of 83% for nucleotides and 90% for amino acids). WCBV was connected to the cluster of GTs 2 and 3. However, although the nucleotide tree joined these genotypes with bootstrap values of 87% (ML) to 92% (NJ), the amino acid tree did not confidently support this relationship: 42% (NJ) to 60% (MP) bootstrap values.

3.2. P gene

The P gene transcriptional start signal for Irkut virus and WCBV was similar to other lyssaviruses: -AACACCYCT-. The coding region of the P gene of Irkut virus consisted of 894 nt and coded for 298 amino acids; the same length found for the P gene of GTs 4 and 5. The coding region of the P gene of WCBV consisted of 891 nt and coded 297 amino acids, like most representatives of GT 1, 6 and 7 (Nadin-Davis et al., 2002). Methionine at positions 53 and 83 of the phosphoprotein was conserved in Irkut virus, but in WCBV it was substituted at these positions by valine and tyrosine, respectively. The WCBV had a double proline at positions 134–135, 137–138 and an additional proline at position 140. The amino acid insertion at 140 is novel to all comparable lyssavirus sequences and produced the first gap in the align-

Table 3
The N and G ectodomain nucleotide and deduced amino acid sequence identities (%) between lyssaviruses of earlier determined phylogroups (Badrane et al., 2001) and WCBV

Compared virus groups	N nucleotides	N amino acids	G ectodomain nucleotides	G ectodomain amino acids
Within phylogroup 1[a]	79.2 ± 0.8 [b]	90.9 ± 0.5	75.5 ± 0.8	83.3 ± 0.8
Within phylogroup 2	83.5 ± 4.4	91.8 ± 2.4	73.9 ± 4.1	81.8 ± 4.8
Between phylogrops 1 and 2	72.2 ± 0.3	81.5 ± 0.4	62.8 ± 0.2	63.1 ± 0.3
Between WCBV and phylogroup 1	72.5 ± 0.3	81.4 ± 0.5	58.3 ± 0.4	54.8 ± 0.3
Between WCBV and phylogroup 2	72.8 ± 1.0	81.5 ± 1.4	60.4 ± 0.6	57.6 ± 0.7

[a] Phylogroup 1 includes genotypes 1, 4-7, Aravan, Khujand and Irkut viruses; phylogroup 2 includes genotypes 2 and 3.
[b] Mean and 95% confidence limit.

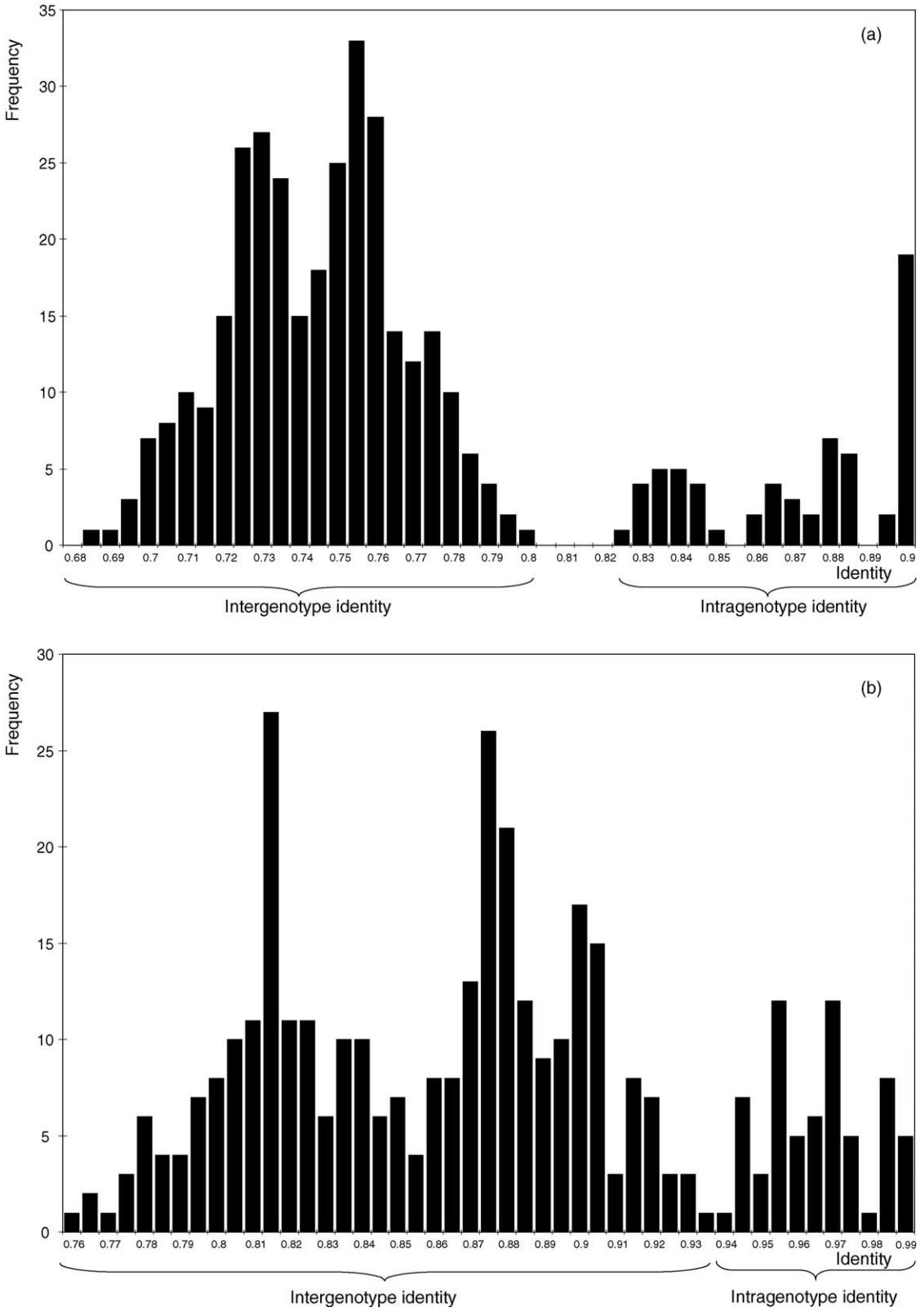

Fig. 1. Distribution of frequences of pairwise identities between the entire N gene sequences of lyssaviruses used in this study (a) nucleotides and (b) amino acids.

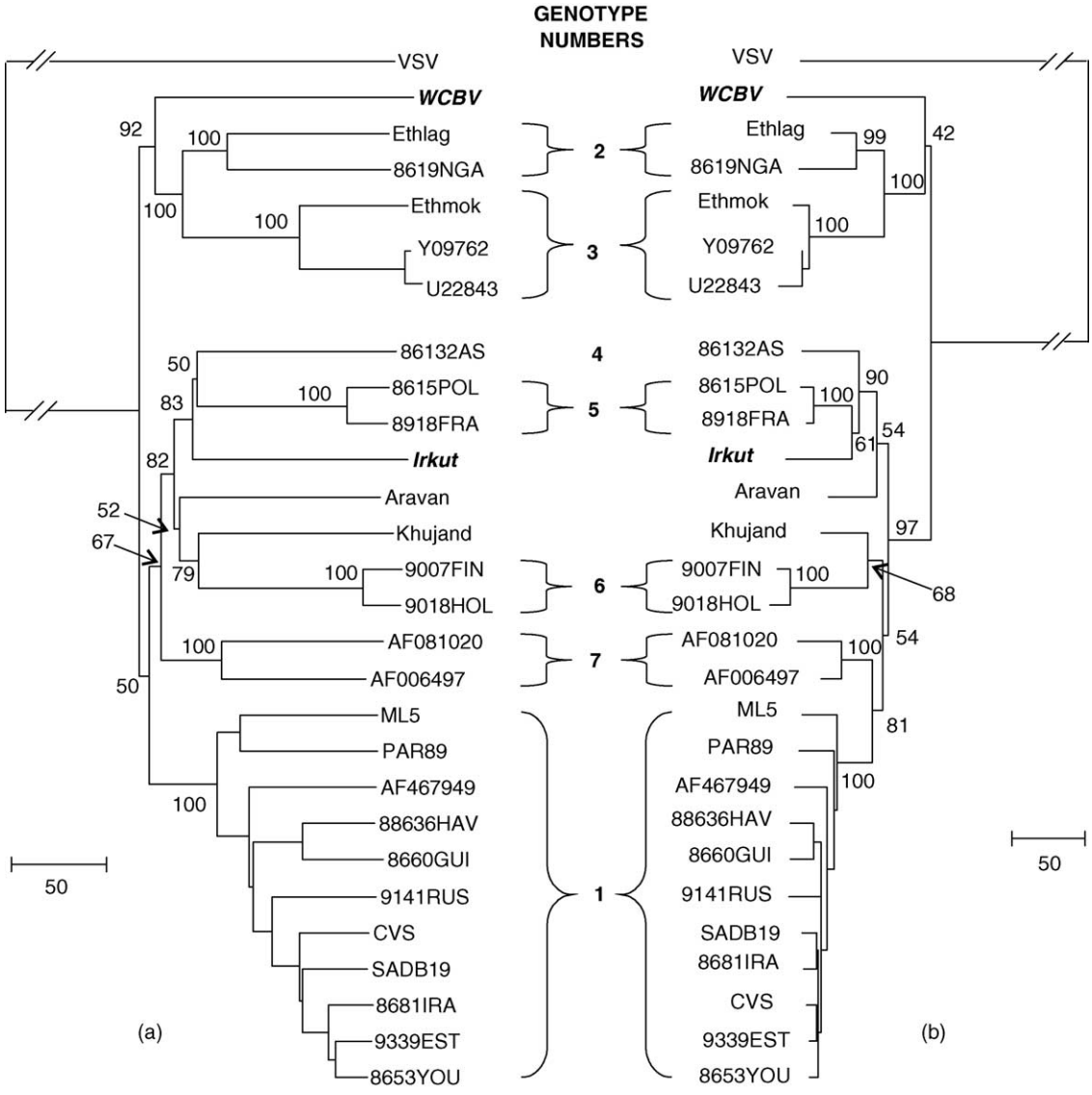

Fig. 2. Phylogenetic trees based on the entire N gene of the *Lyssavirus* genus (a) alignment of 1353 nucleotides, (b) alignment of 451 amino acids; pairwise deletion applied for missed values) obtained by neighbor-joining method. GenBank accession numbers are given when virus names were unavailable. Bootstrap values are presented for key nodes and branch lengths are drawn to scale.

ments. Further gaps occurred in the variable region 471–573 (amino acid positions 159–191) of the P. The critical positions of the binding site for the cytoplasmic light chain of dynein, LC8 (Poisson et al., 2001), were conserved in Irkut virus, but WCBV retained only D_{144} and Q_{148} (Fig. 3). A particularly notable substitution in this site was A_{149} instead of T_{149}, which is present in all other compared sequences. Another stabile motif thought to be an important component of the C-terminal nucleoprotein binding domain of the phosphoprotein, -FSKKYKF- at positions 209–215 (Jacob et al., 2001; Nadin-Davis et al., 2002), was conserved in Irkut virus, and had only one initial substitution (F to Y) in the WCBV phosphoprotein.

Nucleotide and amino acid identities for the P gene are shown in Table 4. Minimum identity of P gene sequences of GT 1 isolates was 75.0 and 80.0% for the nucleotide and amino acid sequences, respectively. The diversity was even more significant between two compared sequences of GT 2: 74.0 and 79.3% for the nucleotide and amino acid sequences, respectively. Irkut virus demonstrated maximum relatedness to GT 5 (71.4 and 71.1% for nucleotide and amino acid sequences, respectively). WCBV was equally different from all other GTs.

The distribution of frequencies of pairwise identities between the P gene sequences revealed three frequency groups (not shown). One of them consisted of identities between the most related sequences (>72% for nucleotide identity, and >78% for amino acid identity). The second group consisted of identities between less related sequences, 60–71% of nucleotide and 56–75% of amino acid identity. The third group consisted of identities between most distant sequence pairs, 55% and less of nucleotide identity, and 49% and less of

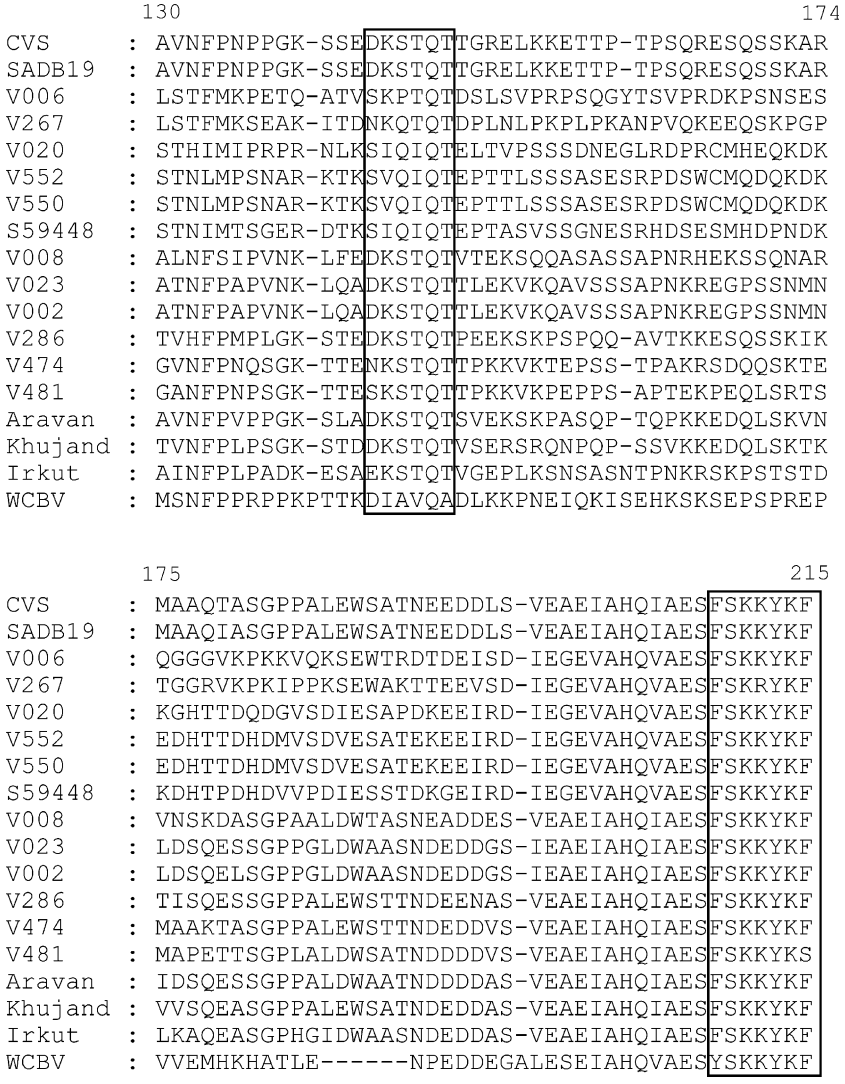

Fig. 3. Partial alignment of lyssavirus phosphoproteins. Amino acid positions are shown above the alignment. LC8 site (aa 144–148) and C-terminal nucleoprotein binding domain (aa 209–215) are framed.

amino acid identity. Aravan and Khujand viruses belonged to the first group since their identity to each other (74.6% for nucleotides and 80.0% for amino acids) and identity of Khujand virus to GT 6 (74.1% for nucleotides and 78.7% for amino acids) were more significant than the intrinsic identity of GT 2. Hence, this separation was different from the separation provided by N gene identities where all these viruses were recognized as representatives of different genotypes.

Table 4
Identity of the P gene and deduced phosphoprotein sequences of lyssaviruses presented in Fig. 4

Genotypes	GT 1	GT 2	GT 3	GT 4	GT 5	GT 6	GT 7	Aravan	Khujand	Irkut	WCBV
GT 1	$\frac{80.0-97.0}{75.0-98.4}$ [a]	44.0–46.3	45.3–47.0	54.0–58.8	58.1–64.4	64.9–68.5	64.5–74.3	63.0–66.6	65.1–69.3	60.9–64.7	41.3–44.3
GT 2	51.5–54.0	$\frac{79.3}{74.0}$	56.3–59.6	43.9–44.5	44.8–46.1	45.3–46.3	43.4–44.6	43.5–43.9	44.7–45.6	44.9–45.9	44.4–44.7
GT 3	49.7–53.7	61.4–63.6	$\frac{81.1-99.3}{79.4-99.3}$	42.2–43.8	43.8–45.8	45.8–47.1	44.2–47.5	44.5–45.5	45.8–47.5	46.2–48.5	42.7–45.7
GT 4	58.7–63.1	53.0	49.6–51.0	$\frac{97.3}{97.9}$	68.3–70.9	57.1–57.8	55.5–56.5	60.2–61.2	60.3–61.0	66.1–66.7	42.3–43.0
GT 5	60.4–63.0	54.6–55.2	50.7–52.0	68.3–69.1	$\frac{98.3}{98.4}$	66.0	59.8–61.1	69.3–70.3	66.2–66.8	70.8–71.1	44.0
GT 6	64.5–65.6	52.5	51.7–52.9	62.9–63.8	66.0		67.5–69.1	78.4	78.7	66.7	43.7
GT 7	66.5–70.2	51.7–54.3	50.6–52.9	59.8–62.2	63.9–64.5	65.9–66.7	$\frac{82.2-98.6}{79.7-99.1}$	66.0–67.3	67.1–69.0	63.4–64.4	42.0–44.7
Aravan	63.3–65.9	50.7–52.0	50.0–52.3	63.3–64.2	68.4–69.0	72.5	67.2–67.7		80.0	67.1	43.3
Khujand	64.8–67.6	51.6–52.6	52.3–54.1	62.8–63.2	66.2	74.1	67.3	74.6		68.4	42.0
Irkut	61.1–65.1	53.1–54.5	52.8–54.4	67.2–67.3	71.4	66.4	64.1–65.2	65.5	66.5		43.7
WCBV	50.3–52.0	51.2	50.3–52.1	51.7–52.0	49.1–49.7	52.6	49.5–50.2	51.2	52.4	51.9	

[a] Upper triangle and above the dash – amino acid identity (shaded); lower triangle and below the dash – nucleotide identity. When only one sequence of a certain genotype was available for comparison, there is no value.

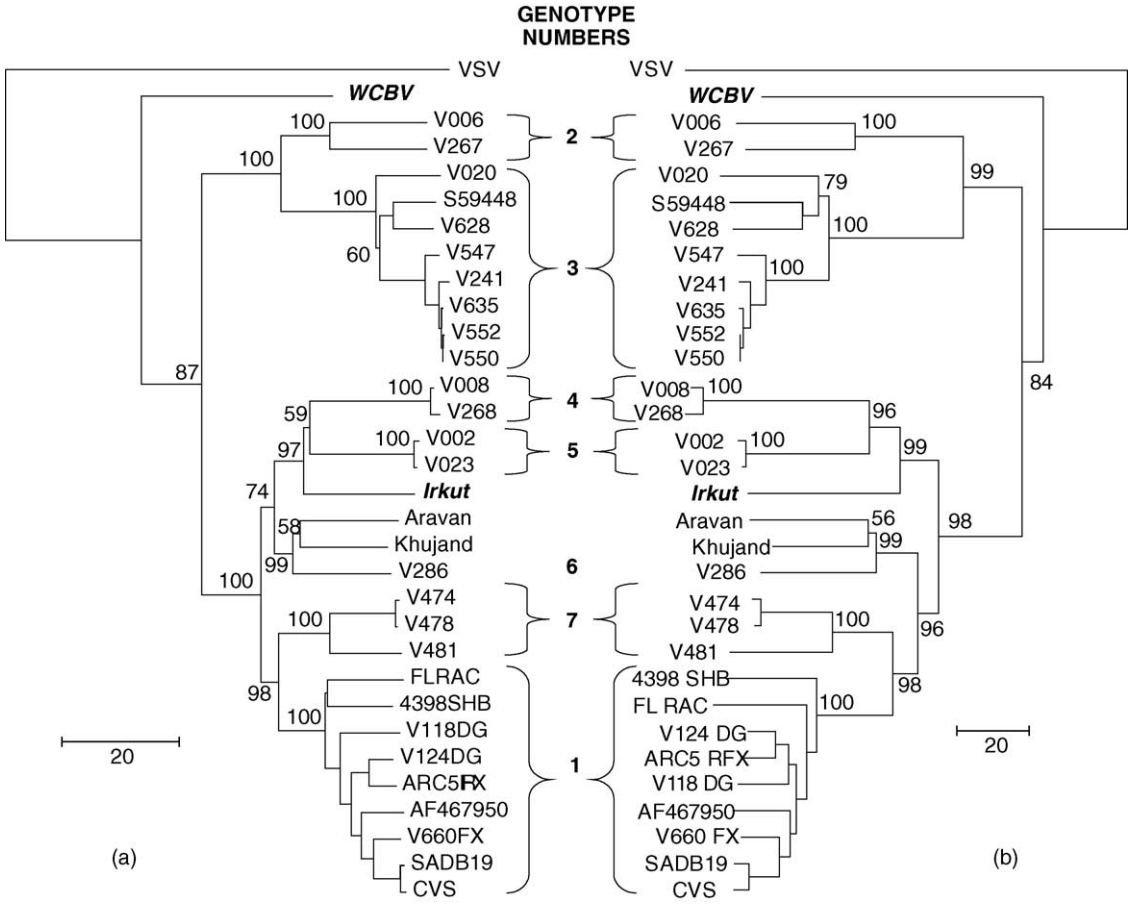

Fig. 4. Phylogenetic trees based on the entire P gene of the *Lyssavirus* genus (a) alignment of 915 nucleotides, and (b) alignment of 305 amino acids; pairwise deletion applied for missed values) obtained by neighbor-joining method. GenBank accession numbers are given when virus names were unavailable. Bootstrap values are presented for key nodes and branch lengths are drawn to scale.

In all phylogenetic constructions based on the P gene, Irkut virus was joined to the cluster of GTs 4 and 5 (Fig. 4). WCBV could not be joined to any particular group and was placed externally to all other lyssaviruses.

3.3. G gene

The G gene of both Irkut and WCBV began with the signal peptide of 19 amino acids common to all lyssaviruses. The entire G gene of Irkut virus coded for 505 amino acids, similar to the G gene of GTs 1, 5 and 6. The G of WCBV coded for 506 amino acids. This is different from all other lyssavirus representatives (Wunner, 2002). The ectodomain of both Irkut virus and WCBV consisted of 439 amino acids, and did not produce any gaps in an alignment with other lyssaviruses, as was also true for the cytoplasmic domain: all alignment gaps were produced along the endodomain only. Along the ectodomain, the recognized antigenic sites were not strictly conserved amongst the different GTs. Only C_{35} and K_{198} (antigenic site II), W_{335} (antigenic site III) and -KG-$_{342\text{-}343}$ (antigenic site a) were conserved in all compared sequences. K_{330} was conserved in Irkut ectodomain but replaced by I_{330} for WCBV. However, WCBV retained K_{331} similarly to some GT 2 representatives (Badrane et al., 2001). R_{333}, which was reported as an important factor of pathogenicity in a peripheral challenge (Dietzschold et al., 1983; Badrane et al., 2001), was conserved in Irkut ectodomain, but not in WCBV, where it was substituted by E_{333}. A potential glycosylation site was found in the WCBV ectodomain at position 334 as it had been published earlier for GT 2 (Badrane et al., 2001). A palmitoylation site (C461) was represented in both viruses as well as a conserved endodomain region -SWESYK- at positions 493–498 (Badrane et al., 2001).

Nucleotide and amino acid identities for the G gene are shown in Table 5. The greatest amount of intragenotype diversity was detected in GT 2: 76.2 and 83.9% identity for nucleotide amino acid sequences respectively. Maximum intergenotype homology was registered between GTs 5 and 6 (72.9 and 80.7% for nucleotides and amino acids, respectively). Irkut virus demonstrated maximum identity with GT 5: 73.5 and 80.0% of nucleotide and amino acid identity, respectively. WCBV was the most diverse member of the identity table, demonstrating overall identity of only 53.3–57.6% at the nucleotide level and 48.1–53.1% identity at the amino acid level, to other GTs.

Table 5
Identity of the G gene and deduced glycoprotein sequences of lyssaviruses presented in Fig. 6

Genotypes	GT 1	GT 2	GT 3	GT 4	GT 5	GT 6	GT 7	Aravan	Khujand	Irkut	WCBV
GT 1	$\frac{84.3-98.4}{78.9-97.7}$ [a]	56.2–57.8	57.4–58.9	67.9–69.6	70.0–71.7	71.5–74.4	73.2–78.2	72.7–74.6	73.7–77.4	66.4–69.0	48.7–50.4
GT 2	57.6–59.2	$\frac{83.9}{76.2}$	72.7–73.7	54.9–56.2	56.8–58.0	55.1–56.4	57.4–58.0	57.5	56.4–57.6	58.7–59.1	52.3–53.1
GT 3	57.3–59.3	68.3–70.1		55.8	57.3	55.1–56.4	57.0–58.2	56.6	56.6	58.0	52.7
GT 4	64.3–65.1	57.3–58.4	58.3	$\frac{97.7}{98.1}$	78.4–78.8	77.0–79.0	71.1–73.6	76.7–77.4	76.3–77.0	72.4–73.7	48.1–48.6
GT 5	65.6–67.2	58.5–60.9	59.8	71.8–72.4	$\frac{97.7}{95.4}$	79.0–80.7	73.2–76.3	82.6–83.0	79.0	80	50.1
GT 6	68.4–70.6	60.1–60.9	58.5–59.6	70.5–72.1	72.3–72.9	$\frac{96.5}{94.0}$	75.5–79.9	83.8	85.9–86.8	75.3–76.5	49.1–50.0
GT 7	69.1–71.6	58.1–59.9	58.5–59.6	66.8–68.4	68.3–69.8	70.9–72.6	$\frac{90.0-99.8}{82.8-99.8}$	76.9–79.5	78.2–82.4	71.4–73.3	50.8–51.2
Aravan	67.9–71.1	59.5–60.1	58.0	69.2–69.5	72.6	74.6	70.3–71.4		83.5	77.3	50.6
Khujand	69.0–69.8	59.1–59.8	58.4	70.7	71.5	77.6–78.2	73.1	74.5		76.2	50.4
Irkut	64.7–66.7	60.1–61.2	59.2	68.1–68.9	73.5	70.2–70.6	65.5–68.5	69.7	69.7		50.2
WCBV	53.3–55.6	56.5–56.9	57.6	55.5	55.2	55.3–56.9	54.9–55.2	55.6	55.3	55.6	

[a] Upper triangle and above the dash – amino acid identity (shaded); lower triangle and below the dash – nucleotide identity. When only one sequence of a certain genotype was available for comparison, there is no value.

The distribution of frequencies of the pairwise identities between the G gene sequences, and particularly ectodomain amino acid sequences (Fig. 5), demonstrated multiple overlaps between Aravan and Khujand viruses and GTs 5–7, so that it was impossible to define a threshold between intragenotype and intergenotype identities using this part of the genome.

When phylogenetic analysis was performed (Fig. 6), Irkut virus was joined to GT 5 by all methods used for either nucleotide or amino acid sequences. However, this placement was not confidently supported, with bootstrap values of only 43–69%. Nonetheless, the position of Irkut virus inside the cluster of GTs 4 and 5 was confirmed. WCBV could not be joined to any cluster and was positioned externally to all lyssaviruses. An identical topology was obtained by different phylogenetic methods applied for the ectodomain nucleotide and amino acid sequences (results not shown).

3.4. Positive selection analysis

Analysis of N and G gene phylogenies revealed no significant evidence of positive selection. For the N gene sequences, models allowing for positive selection were significantly accepted in two cases ($p < 0.0001$), however, no ω value greater than 1 was estimated. Similarly, for the G gene sequences,

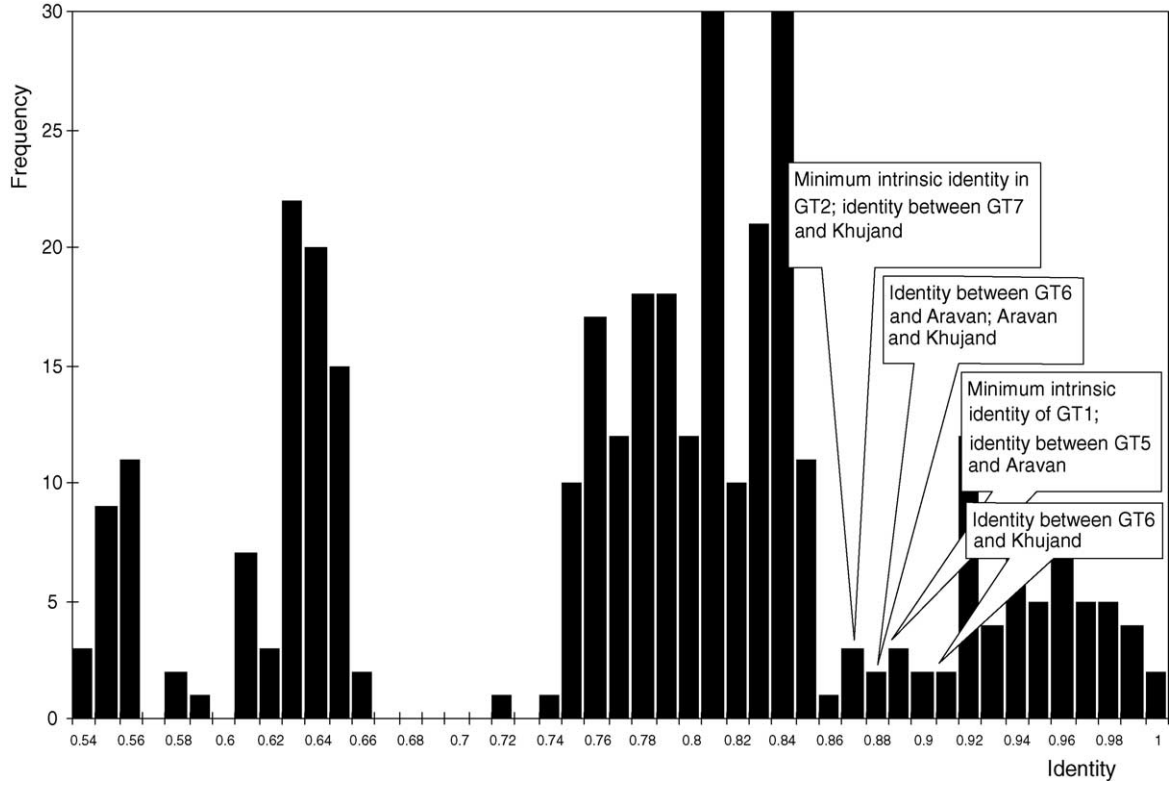

Fig. 5. Distribution of frequencies of pairwise identities between the ectodomain amino acid sequences of lyssaviruses used in the current study.

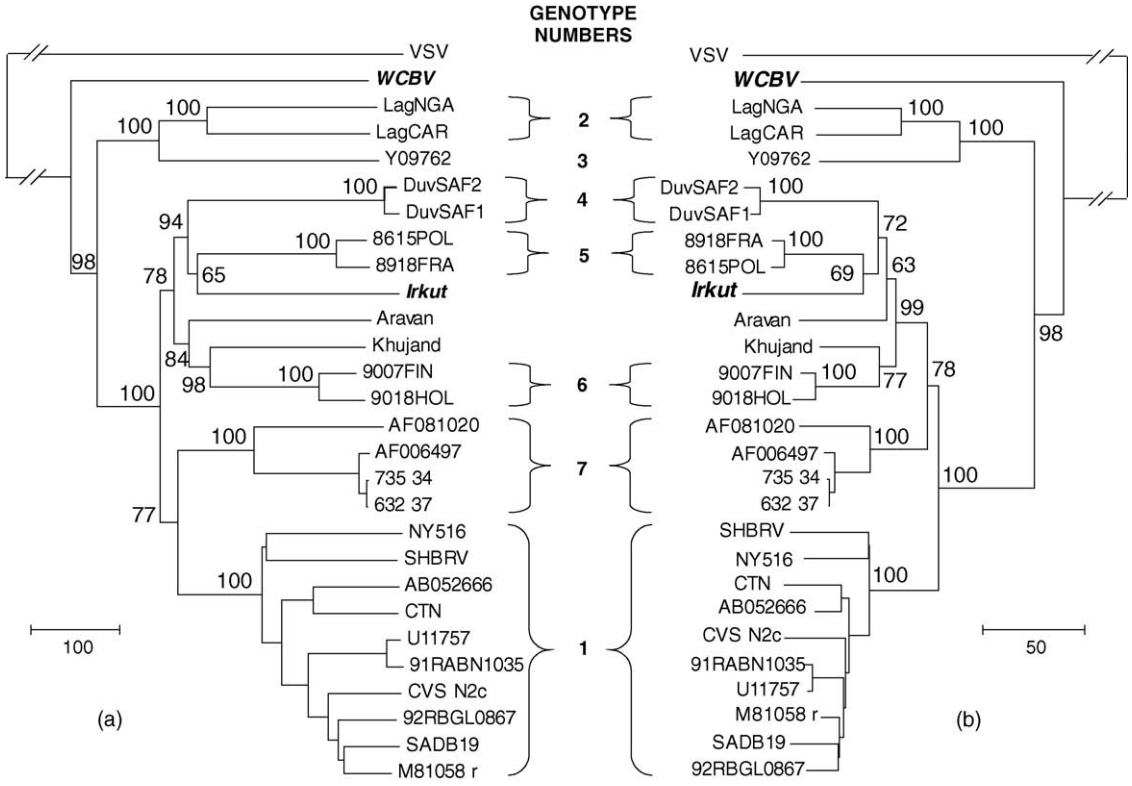

Fig. 6. Phylogenetic trees based on the entire G gene of the *Lyssavirus* genus (a) alignment of 1599 nucleotides, and (b) alignment of 533 amino acids; pairwise deletion applied for missed values) obtained by neighbor-joining method. Genbank accession numbers are given when virus names were unavailable. Bootstrap values are presented for key nodes and branch lengths are drawn to scale.

one null model was rejected ($p < 0.0001$), but no codons were estimated to have a $\omega > 1$.

For P gene sequences, there was very limited evidence of positive selection. One null model (M1 which allows for neutral evolution where $($ is either 1 or 0) was rejected significantly ($p < 0.0001$) in favor of M2, which allows for a $\omega > 1$. The M2 model was the only one of the three positive selection models (together with M3 and M8; Woelk et al., 2002) which estimated a codon class with a $\omega < 1$ (3.18). Consequently, the presence of five positively selected codons along the P gene (amino acid positions 61, 67, 151, 159 and 161 [where position 1 is the conserved ATG start codon]) determined by this model are highly speculative.

4. Discussion

Definition of quantitative criteria for inclusion of a new genotype into the *Lyssavirus* genus became problematic after discovery of Aravan and Khujand viruses (Kuzmin et al., 2003). Criteria proposed earlier were quite acceptable for GTs 1–7 to be determined independently on the gene used: intragenotype identities were greater than intergenotype identities, and clusters of certain genotype members were supported by significant bootstrap values (Bourhy et al., 1993; Tordo et al., 1993; Kissi et al., 1995; Badrane et al., 2001; Johnson et al., 2002). Introducing the newly discovered lyssaviruses required a reassessement of the old approaches.

In this analysis, we obtained different quantitative separations within the genus according to the pairwise identities of the N and P gene. The G gene sequences, either complete or partial could not separate genotypes due to multiple overlaps. We propose that it would be necessary to select one of the genes for taxonomic purposes. But which one provides the best subdivision?

The N gene provides the most clear and unequivocal subdivision of the lyssavirus groups (Fig. 1). It is commonly sequenced as part of molecular epidemiological studies, and historically is associated with earlier lyssavirus classification based on reactivity patterns with anti-nucleocapsid monoclonal antibodies (Bourhy et al., 1993; Kissi et al., 1995). We suggest this gene is the most appropriate for quantitative genotype definition, and the nucleotide sequence should be preferred to the amino acid sequence due to a higher degree of resolution. The results of our study suggest that the threshold of 80% nucleotide identity for sequences of the entire N gene (Kissi et al., 1995) still appears to be reasonable. Moreover, we would raise this threshold to 82%. According to this approach, both Irkut virus and WCBV, as well as the earlier described Aravan and Khujand viruses (Arai et al., 2003; Kuzmin et al., 2003), should be considered as different lyssavirus genotypes. Partial N gene sequences may be used for classification purposes (Amengual et al., 1997; Bourhy et

al., 1999), but their appropriateness must be assessed more thoroughly for recently discovered lyssaviruses.

Use of some qualitative criteria for taxonomy purposes, such as geographic distribution and host species, is an alternative approach. Unfortunately, our present knowledge of the natural history of most non-rabies lyssaviruses is too limited for successful implementation of this suggestion.

All our phylogenetic constructions demonstrate that GTs 4–6, together with Aravan, Khujand and Irkut viruses, present a solid phylogroup of Old World bat lyssaviruses (Fig. 7). We could not provide a straight statistical basis for their joining into one phylogroup as it has been done for phylogroups 1 and 2 (Badrane et al., 2001), due to divergence of marginal members (GTs 4 and 6). Nonethless, the relatedness of these viruses to each other is clearly apparent. This relatedness is seen with each gene, and explains why Aravan and Khujand viruses could not be separated from GT 6 according to the P gene sequence, and why multiple overlaps between members of this phylogroup occurred according to the G gene sequence. However, this is not a rationale to tighten some of these lineages into one genotype if we consider others as independent genotypes. All of these lineages are clearly separated as independent genotypes according to the N gene sequences. Relatedness of these viruses cannot be solely attributed to convergent evolution as a result of host species adaptation: bat lyssaviruses of other GTs do not tend to be particularly related to members of this phylogroup, except GT 7. As was shown earlier (Kuzmin et al., 2003) and confirmed in this study, GT 7 is more related to Khujand virus and GT 6 according to the G amino acid sequence, while according to the N and P sequences it is most related to the GT 1. This particular homology of the G may be suggested as evidence of convergent or parallel evolution.

No significant evidence of positive selection was found in the compared sequences. Previously, evidence of purifying selection has been suggested as a result of genetic constraints on the rabies virus genes analyzed (G and N) or a lack of adequate selection pressure (Badrane and Tordo, 2001; Holmes et al., 2002). We do not consider the results of this study to be unexpected, considering the vast number of independent evolutionary paths that have been pooled within the analysis, a problem that may inhibit tests for positive selection on a genotype basis due to the compartmentalization that undoubtedly governs lyssavirus epidemiology (e.g., Smith, 2002). To correctly assess the role of positive selection in lyssavirus evolution, analysis of phylogenetically distinct lineages may be necessary.

WCBV is the most divergent representative of the *Lyssavirus* genus described to date. The limited relatedness of the N gene sequences of WCBV to GTs 2 and 3 may derive from parental homology, although there are no additional data to support this assertion. The phylogenetic analysis described here raises interesting questions regarding the origin and evolution of the *Lyssavirus* genus. Do Old World bat lyssaviruses originate from Africa, as was proposed earlier for the EBLVs (Shope, 1982; Amengual et al., 1997; Serra-Cobo et al., 2002)? They might, since all putative reservoir species appear to be broadly distributed, due in no small measure to the power of flight. However, only one member of this group, DUVV, is present in Africa. Other lineages are found in Eurasia. Is this observation evidence of independent evolution after an initial introduction of a progenitor virus from Africa? DUVV is no more closely related to LBV and MOKV than it is to EBLV-1, EBLV-2, Aravan, Khujand and Irkut viruses. On the other hand, MOKV, which has been isolated only in Africa was demonstrated to be antigenically related to the African arthropod rhabdoviruses, Obodhiang and Kotonkan, hypothesized as examples for probable progenitors of the lyssavirus genus (Shope, 1982). Has WCBV derived from Africa as well? The principal host of this virus is unknown. It was isolated from *Miniopterus schreibersi*, a broadly distributed colonial species. This bat species does not perform long-distance seasonal migrations, but it is quite mobile and, according to tracking data, the population occupying caves on both slopes of the Caucasian ridge is permanently intermixed (Gazaryan, 1999). Thus, more distant movements appear probable. Additional investigations should be carried out to determine genetic and antigenic relations between WCBV, LBV, MOKV, Obodhiang and Kotonkan viruses, together with uncharacterized rhabdoviruses.

Another important question addresses pathogenicity of WCBV. As had been repeatedly demonstrated, R_{333} (or K_{333}) in the glycoprotein ectodomain, presented in all phylogroup 1 lyssaviruses (except artificially modified laboratory strains), play crucial role in their peripheral pathogenicity (Dietzschold et al., 1983; Badrane et al., 2001; Mebatsion, 2001). Pathogenic importance was also suggested for the LC8 binding site of the phosphoprotein (Poisson et al., 2001; Mebatsion, 2001). The phylogroup 2 members (LBV and MOKV) have D_{333} in the glycoprotein, and the LC8 binding site of their phosphoprotein is different from that of the phylogroup 1 lyssaviruses. LBV and MOKV were apathogenic for mice and dogs, and of low pathogenicity for monkeys by the peripheral route (Tignor et al., 1973; Dietzschold et al., 1983; Badrane et al., 2001). WCBV has E_{333} in the glycoprotein ectodomain and the LC8 binding site of the phosphoprotein is different from either 1 or 2 phylogroup lyssaviruses. In our experiments WCBV was apathogenic for 3-week-old mice by intramuscular and oral routes even when 10^6 mouse intracerebral lethal doses were given. But syrian hamsters challenged intramuscularly with the same virus dose developed typical rabies symptoms and all succumbed (100% mortality). Also, we should not forget that WCBV and phylogroup 2 lyssaviruses are not laboratory strains. They were isolated from a number of naturally infected mammalian species (MOKV was also isolated from humans), so there should be some adaptive mechanisms providing their natural circulation.

The public health implications of newly discovered lyssaviruses require introspection. Although no human infection caused by viruses other then GTs 1, 5 and 6 have been reported from this region, such observations may be as

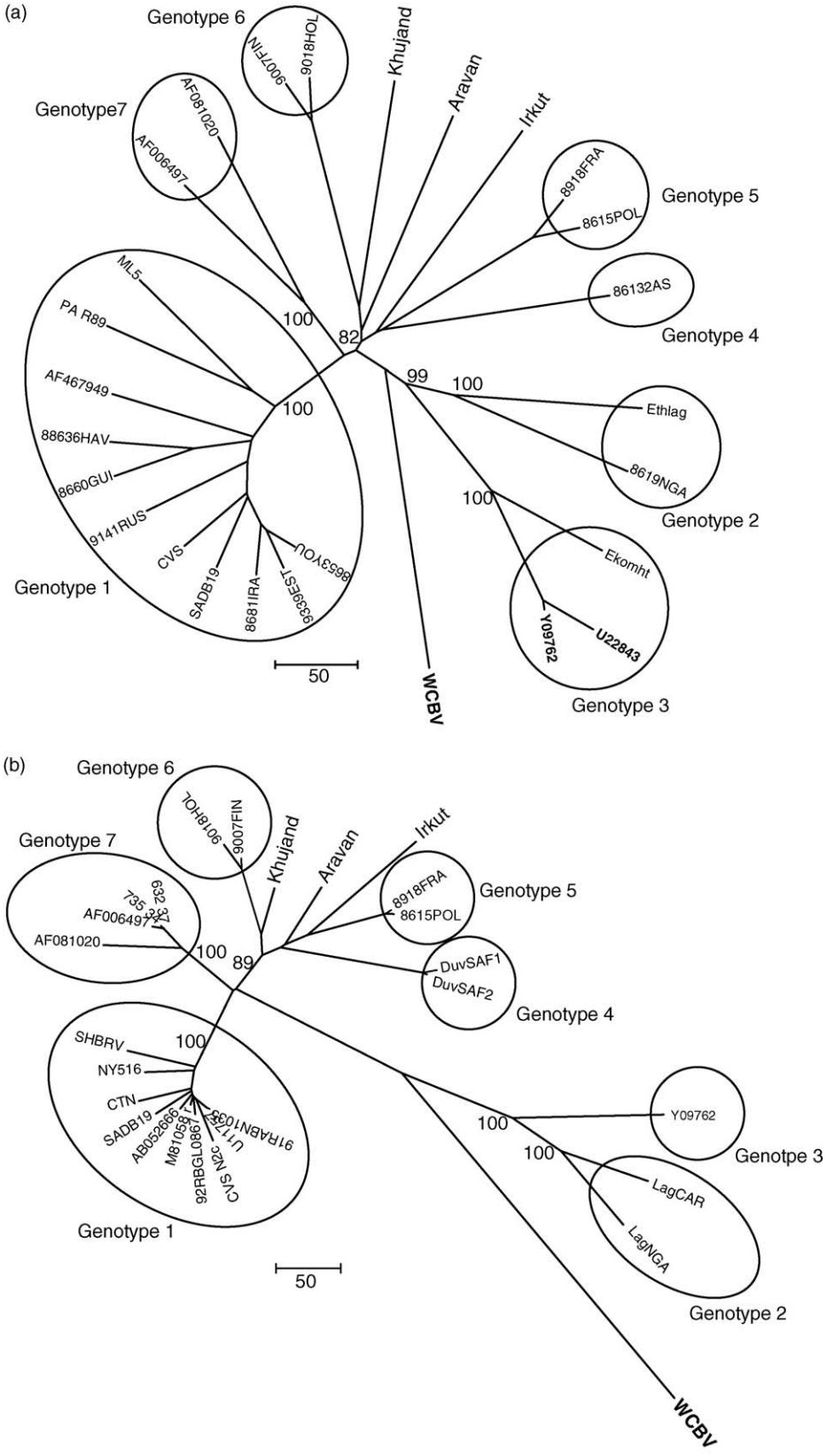

Fig. 7. Unrooted phylogenetic trees based on the entire lyssavirus N gene nucleotide sequences (a) and glycoprotein ectodomain amino acid sequences (b). The genotypes (GT) recognized earlier are enclosed in circles. Genbank accession numbers are given when virus names are unavailable. Bootstrap values are presented for key nodes and branch lengths are drawn to scale.

a result of inadequate diagnostic tests or inappropriate epidemiological surveillance (Kuzmin et al., 2003). Moreover, considering bat mobility and the potential for spatial spread of infection, the application of the term "rabies-free" should be used conservatively. For example, the Irkutsk region has been considered free of rabies for 35 years before the isolation of Irkut virus (Botvinkin et al., 2003). Similarly, the United Kingdom was believed to be "rabies-free" until 1996, when EBLV-2 was isolated from bats (Whitby et al., 1996), and a human case of this infection was registered in 2002 (Fooks et al., 2002). Public health awareness should be raised, and attention paid to the protective efficacy of commercially available rabies vaccines, as well as immune globulin reactivity, with these non-rabies lyssaviruses. Anti-rabies biologicals provide incomplete protection of experimental animals against EBLVs, limited protection against LBV, and no protection against MOKV (Fekadu et al., 1988; Lafon et al., 1988; Mebatsion et al., 1992; Bahlou et al., 1998; Jallet et al., 1999; Nel et al., 2003). We should expect the same lack of efficacy against WCBV.

Acknowledgements

We are grateful to Drs. T. Mebatsion and J.H. Cox, (Federal Institute for Animal Viral Diseases, Tubingen, Germany) for the N gene sequences of LBV and MOKV viruses originating from Ethiopia. We thank J.S. Smith (CDC, Atlanta, GA, USA) for valuable discussions during the study. IVK was supported by the Russian Foundation of Basic Research (grant 00-04-48004) and Association of Public Health Laboratories, international fellowship program of 2002–2003. GJH was funded by an ASM/NCID post-doctoral fellowship.

References

Amengual, B., Whitby, J.E., King, A., Serra Cobo, J., Bourhy, H., 1997. Evolution of European bat lyssaviruses. J. Gen. Virol. 78, 2319–2328.

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. Inf. Dis. 9, 333–337.

Arguin, P.M., Murray-Lillibridge, K., Miranda, M.E., Smith, J.S., Calaor, A.B., Rupprecht, C.E., 2002. Serologic evidence of Lyssavirus infections among bats, the Philippines. Emerg. Inf. Dis. 8, 258–262.

Badrane, H., Bahloul, C., Perrin, P., Tordo, N., 2001. Evidence of two Lyssavirus phylogroups with distinct pathogenecity and immunogenecity. J. Virol. 75, 3268–3276.

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

Bahlou, C., Jacob, Y., Tordo, N., Perrin, P., 1998. DNA-based immunization for exploring the enlargement of immunological cross-reactivity against the lyssaviruses. Vaccine 16, 417–425.

Benmansour, A., Brahmi, M., Tuffereau, C., Coulon, P., Lafay, F., Flamand, A., 1992. Rapid sequence evolution of street rabies glycoprotein is related to the highly heterogeneous nature of the viral population. Virology 187, 33–45.

Botvinkin, A.D., Poleschuk, E.M., Kuzmin, I.V., Borisova, T.I., Gazaryan, S.V., Yager, P., Rupprecht, C.E., 2003. Novel lyssavirus isolated from bats Russia. Emerg. Inf. Dis. 9, 1623–1625.

Boulger, I.R., Porterfield, J.S., 1958. Isolation of a virus from Nigerian fruit bats. Trans. Roy. Soc. Trop. Med. Hyg. 52, 421–424.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J.F., Holmes, E.C., 1999. Ecology and evolution of rabies virus in Europe. J. Gen. Virol. 80, 2545–2557.

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

Conzelmann, K.K., Cox, J.H., Schneider, L.G., Thiel, H.J., 1990. Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. Virology 175, 485–499.

da Cruz, F.W., McBride, A.J.A., Conceição, F.R., Dale, J.W., McFadden, J., Dellagostin, O.A., 2002. Expression of the B-cell and T-cell epitopes of the rabies virus nucleoprotein in *Mycobacterium bovis* BCG and induction of an humoral response in mice. Vaccine 20, 731–736.

Dietzschold, B., Wunner, W.H., Wiktor, T.J., Lopes, A.D., Lafon, M., Smith, C.L., Koprowski, H., 1983. Characterization of an antigenic determinant of the glycoprotein that correlates with pathogenicity of rabies virus. Proc. Natl. Acad. Sci. U.S.A. 80, 70–74.

Dietzschold, B., Lafon, M., Wang, H., Otvos, L.J., Celis, E., Wunner, W.H., Koprowski, H., 1987. Localization and immunological characterization of antigenic domains of the rabies virus internal N and NS proteins. Virus Res. 8, 103–125.

Ertl, H., Dietzchold, B., Otvos, L., 1991. T helper cell epitope of rabies virus nucleoprotein defined by tri- and tetrapeptides. Eur. J. Immunol. 21, 1–10.

Fekadu, M., Shaddock, J.H., Sanderlin, D.W., Smith, J.S., 1988. Efficacy of rabies vaccines against Duvenhage virus isolated from European house bats (*Eptesicus serotinus*), classic rabies and rabies-related viruses. Vaccine 6, 533–539.

Felsenstein, J., 1993. PHYLIP: Phylogenetic Inference Package (Version 3.5c). Department of Genetics, University of Washington, Seattle, WA, USA.

Foggin, C.M., 1988. Rabies and rabies-related viruses in Zimbabwe: historical, virological and ecological aspects. Thesis, University of Zimbabwe, Harare.

Fooks, A.R., Finnegan, C., Johnson, N., Mansfield, K., McElhinney, L. Manser, P., 2002. Human case of EBL type 2 following exposure to bats in Angus, Scotland. Vet. Rec. 151, 679.

Fu, Z.F., Wunner, W.H., Dietzschold, B., 1994. Immunoprotection by rabies virus nucleoprotein. In: Rupprecht, C.E., Dietzschold, B., Koprowski, H. (Eds.), Lyssaviruses. Springer, Berlin, Germany, pp. 207–218.

Fraser, C.F., Hooper, P.T., Lunt, R.A., Gould, A.R., Glesson, L.J., Hyatt, A.D., Russell, G.M., Kattenbelt, J.A., 1996. Encephalitis caused by a Lyssavirus in fruit bats in Australia. Emerg. Inf. Dis. 2, 327–331.

Gazaryan, S.V., 1999. New data on occurrence of Schreiber's bats (*Miniopterus schreibersi*) in the Western Caucasus. Plecotus et al. 2, 88–93.

Gould, A.R., Kattenbelt, J.A., Gumley, S.G., Lunt, R.A., 2002. Characterization of an Australian bat lyssavirus variant isolated from an insectivorous bat. Virus Res. 89, 1–28.

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.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95–98.

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

Jacob, Y., Real, E., Tordo, N., 2001. Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. J. Virol. 75, 9613–9622.

Jallet, C., Jacob, Y., Bahloul, C., Drings, A., Desmezières, E., Tordo, N., Perrin, P., 1999. Chimeric lyssavirus glycoproteins with increased immunological potential. J. Virol. 73, 225–233.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. Trends. Biochem. Sci. 23, 403–405.

Johnson, N., McElhinney, L.M., Smith, J., Lowings, P., Fooks, A.R., 2002. Phylogenetic comparison of the genus Lyssavirus using distal coding sequences of the glycoprotein and nucleoprotein genes. Arch. Virol. 147, 2111–2123.

King, A., Crick, J., 1988. Rabies-related viruses. In: Campbell, J.B., Charlton, K.M. (Eds.), Rabies. Kluwer Acad. Publisher, Boston, pp. 178–199.

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

Koprowski, H., 1996. The mouse inoculation test. In: Meslin F.-X., Kaplan, M.M., Koprowski, H. (Eds.), Laboratory Techniques in Rabies, 4th ed. WHO, Geneva, pp. 80–86.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

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.

Kuzmin, I.V., Botvinkin, A.D., Khabilov, T.K., 2002. A lyssavirus was isolated from a whiskered bat in Northern Tajikistan. Plecotus et al. 4, 75–81, in Russian.

Kuzmin, I.V., Botvinkin, A.D., Rybin, S.N., Bayaliev, A.B., 1992. A lyssavirus with unusual antigenic structure isolated from a bat in Southern Kyrghyzstan. Voprosy Virusologii 5–6, 256–259 (in Russian).

Lafon, M., Bourhy, H., Sureau, P., 1988. Immunity against the European bat rabies (Duvenhage) virus induced by rabies vaccines: an experimental study in mice. Vaccine 6, 362–368.

Larson, J.K., Wunner, W.H., 1990. Nucleotide and deduced amino acid sequences of the nominal nonstructural phosphoprotein of the ERA, PM and CVS-11 strains of rabies virus. Nucl. Acids Res. 18, 7172.

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.

Lumio, J., Hillbom, M., Roine, R., Ketonen, L., Haltia, M., Valle, M., Neuvonen, E., Lahdeviktra, J., 1986. Human rabies of bat origin in Europe. Lancet 1, 378.

Mebatsion, T., Cox, J.H., Frost, J.W., 1992. Isolation and characterization of 115 street rabies virus isolates from Ethiopia by using monoclonal antibodies: identification of 2 isolates as Mokola and Lagos bat viruses. J. Inf. Dis. 166, 972–977.

Mebatsion, T., 2001. Extensive attenuation of rabies virus by simultaneously modifying the dynein light chain binding site in the P protein and replacing Arg333 in the G protein. J. Virol., 11496–11502.

Meredith, C.D., 1980. Rabies confirmed in Megachiroptera in South Africa. Rabies Inf. Exchange 3, 26.

Meredith, C.D., Rossouw, A.R., Van Praag Koch, H., 1971. An unusual case of human rabies thought to be of Chiropteran origin. S. Afr. Med. J. 45, 767–769.

Morimoto, K., Patel, M., Corisdeo, S., Hooper, D.C., Fu, Z.F., Rupprecht, C.E., Koprowski, H., Dietzschold, B., 1996. Characterization of a unique variant of bat rabies virus responsible for newly emerging human cases in North America. Proc. Natl. Acad. Sci. U.S.A. 93, 5653–5658.

Morimoto, K., Hooper, D.C., Carbaugh, H., Fu, Z.F., Koprowski, H., Dietzschold, B., 1998. Rabies virus quasispecies: implications for pathogenesis. Proc. Natl. Acad. Sci. U.S.A. 95, 3152–3156.

Nadin-Davis, S.A., Huang, W., Wandeler, A.I., 1996. The design of strain-specific polymerase chain reactions for discrimination of the raccoon rabies virus strain from indigenous rabies viruses of Ontario. J. Virol. Methods 57, 1–14.

Nadin-Davis, S.A., Sampath, M.I., Casey, G.A., Tinline, R.R., Wandeler, A.I., 1999. Phylogeographic patterns exhibited by Ontario rabies virus variants. Epidemiol. Inf. 123, 325–336.

Nadin-Davis, S.A., Huang, W., Armstrong, J., Casey, G.A., Bahloul, C., Tordo, N., Wandeler, A.I., 2001. Antigenic and genetic divergence of rabies viruses from bat species indigenous to Canada. Virus Res. 74, 139–156.

Nadin-Davis, S.A., Abdel-Malik, M., Armstrong, J., Wandeler, A., 2002. Lyssavirus P gene characterization provides insight into the phylogeny of the genus and identifies structural similarities and diversity within the encoded phosphoprotein. Virology 298, 286–305.

Nel, L., Jacobs, J., Jafta, J., Von Teichman, B., Bingham, J., 2000. New cases of Mokola virus infection in South Africa: a genotypic comparison of Southern African virus isolates. Virus Genes 20, 103–106.

Nel, L.H., Niezgoda, M., Hanlon, C.A., Morril, P.A., Yager, P.A., Rupprecht, C.E., 2003. A comparison of DNA vaccines for the rabies-related virus, Mokola. Vaccine 21, 2598–2606.

Pal, S.R., Arora, B.M., Chuttani, P.N., Broer, S., Choudhury, S., Joshi, R.M., 1980. Rabies virus infection of a flying fox bat, *Pteropus poliocephalus*, in Chandigarh, Northern India. Trop. Geogr. Med. 32, 265–267.

Poisson, N., Real, E., Gaudin, Y., Vaney, M.C., King, S., Jacob, Y., Tordo, N., Blondel, D., 2001. Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P. J. Gen. Virol. 82, 2691–2696.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Sacramento, D., Bourhy, H., Tordo, N., 1991. PCR technique as an alternative method for diagnosis and molecular epidemiology of rabies virus. Molec. Cell. Prob. 5, 229–240.

Schneider, L.G., Barnard, B.J.H., Schneider, H.P., 1985. Application of monoclonal antibodies for epidemiological investigations and oral vaccination studies. 1. African viruses. In: Kuwert, E., Merieux, C., Koprowski, H., Bogel, K. (Eds.), Rabies in the Tropics. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, pp. 47–59.

Selimov, M.A., Tatarov, A.G., Botvinkin, A.D., Klueva, E.V., Kulikova, L.G., Khismatullina, N.A., 1989. Rabies-related Yuli-virus; identification with a panel of monoclonal antibodies. Acta Virol. 33, 542–545.

Serra-Cobo, J., Amengual, B., Abellan, C., Bourhy, H., 2002. European bat lyssavirus infection in Spanish bat populations. Emerg. Inf. Dis. 8, 414–420.

Shope, R.E., Murphy, F.A., Harrison, A.K., Causey, O.R., Kemp, G.E., Simpson, D.I.H., Moore, D., 1970. Two African viruses serologically and morphologically related to rabies virus. J. Virol. 6, 690–692.

Shope, R.E., 1982. Rabies-related viruses. Yale J. Biol. Med. 55, 271–275.

Smith, J.S., 2002. Molecular epidemiology. In: Jackson, A.C., Wunner, W.H. (Eds.), Rabies. Acad. Press, San Diego, pp. 79–113.

Smith, P.C., Lawhaswasdi, K., Vick, W.E., Stanton, J.S., 1968. Isolation of rabies virus from fruit bats in Thailand. Nature 216, 384.

Sureau, P., Tignor, G.H., Smith, A.L., 1980. Antigenic characterization of the Bangui strain (ANCB-672d) of Lagos bat. Ann. Virol. 131, 25–32.

Swofford, D.L., 2000. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b6. Sinauer Associates, Sunderland, MA.

Tang, Q., Orciari, L.A., Rupprechti, C.E., Zhao, X., 2000. Sequencing and Positional Analysis of the Glycoprotein Gene of Four Chinese Rabies Viruses. Zhongguo Bingduxue 15, 22–33.

Tignor, G.H., Shope, R.E., Bhatt, P.N., 1973. Experimental infection of dogs and monkeys with two rabies serogroup viruses, Lagos Bat and Mokola (IbAn 27377): clinical, serologic, virologic, and fluorescent-antibody studies. J. Inf. Dis. 128, 471–478.

Tordo, N., Poch, O., Ermine, A., Keith, G., Rougeon, F., 1986. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? Proc. Natl. Acad. Sci. U.S.A. 83, 3914–3918.

Tordo, N., Badrane, H., Bourhy, H., Sacramento, D., 1993. Molecular epidemiology of Lyssaviruses: focus on the glycoprotein and pseudogenes. Onderstepoort J. Vet. Res. 60, 315–323.

Van der Merwe, M., 1982. Bats as vectors of rabies. S. Afr. J. Sci. 78, 421–422.

Whitby, J.E., Johnstone, P., Parsons, G., King, A.A., Hutson, A.M., 1996. Ten-year survey of British bats for the existence of rabies. Vet. Rec. 139, 491–493.

Woelk, C.H., Pybus, O.G., Jin, L., Brown, D.W.G., Holmes, E.C., 2002. Increased positive selection pressure in persistent (SSPE) versus acute measles virus infections. J. Gen. Virol. 83, 1419–1430.

Wunner, W.H., 2002. In: Jackson, A.C., Wunner, W.H. (Eds.), Rabies. Acad. Press, San Diego, pp. 23–78.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.