This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.



For guidance, see the notes written in blue and the separate document “Help with completing a taxonomic proposal”

Please try to keep related proposals within a single document.

Part 1: **TITLE, AUTHORS, etc**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Code assigned:** | ***2019.012M*** | | | | (to be completed by ICTV officers) |
| **Short title:** Create five new genera (*Sawgrhavirus, Barhavirus, Zarhavirus, Lostrhavirus, Mousrhavirus*), including eight new species and one previously unassigned species, in the family *Rhabdoviridae* | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| ICTV *Rhabdoviridae* Study Group:  Peter J. Walker  Kim R. Blasdell  Ralf G. Dietzgen  Anthony R. Fooks  Juliana Freitas-Astúa  Hideki Kondo  Gael Kurath  David M. Stone  Robert B. Tesh  Nikos Vasilakis  Anna E. Whitfield | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| Peter J. Walker, [peter.walker@uq.edu.au](mailto:peter.walker@uq.edu.au) | | | | | |
| **List the ICTV study group(s) that have seen this proposal:** | | | | | |
| A list of study groups and contacts is provided at <http://www.ictvonline.org/subcommittees.asp> . If in doubt, contact the appropriate subcommittee chair (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal) | | | ICTV *Rhabdoviridae* Study Group | | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | | | |
| The Study Group considered several options for the assignment of new genera to accommodate the viruses described in this proposal, including:  a) Assign all viruses in the clade described here to a single new genus.  *There are two problems with this approach. Firstly, the viruses differ considerably in ecology with some being tick-borne viruses and others being mosquito-borne viruses. Furthermore, the tick-borne virus clusters and the mosquito-borne viruses are not monophyletic. Secondly, two clusters of these viruses are recognised as two different serogroups (Sawgrass serogroup and Bahia Grande serogroup) based on cross-reactions in neutralisation and/or complement fixation tests. In general, rhabdoviruses of different serogroups have been assigned to different genera.*  b) Create only two new genera (*Sawgrhavirus* and *Barhavirus*) and two new unassigned species.  *ICTV requires now that there be no new unassigned species.*  c) Create only two new genera (*Sawgrhavirus* and *Barhavirus*) and delay the classification of ZARV and LSTRV until more information is available.  *If this approach was applied more broadly, many rhabdoviruses would likely remain unclassified. The option was preferred by only 1 of the 9 responding Study Group members.*  d) Create two new genera (*Sawgrhavirus* and *Barhavirus*) and another single new genus containing all three other viruses.  *The proposed third genus would include both tick-borne viruses (ZARV and LSTRV) and a mosquito-borne virus (MOUV) with sequence divergence levels that reflect genus-level rather than species-level divergence.*  e) Create five new genera (*Sawgrhavirus*, *Barhavirus*, *Zarhavirus*, *Lostrhavirus*, *Mousrhavirus*), 8 new species and one reassigned species.  *Although three of the new genera would contain only a single species, this was the approach preferred by 10 of the 11 responding Study Group members.*  Based on these considerations, option e) is the approach that has been adopted in this proposal. | | | | | |
|  | | | | | |
| Date first submitted to ICTV: | | | | 25 January 2019 | |
| Date of this revision (if different to above): | | | |  | |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
|  |

**Part 2**: **PROPOSED TAXONOMY**

|  |
| --- |
| Present the proposed new taxonomy on accompanying spreadsheet |
| **Name of accompanying spreadsheet:** 2018.012M.A.v1.Rhabdoviridae\_5gen8sp1reasp.xlsx |

**Part 4:** **APPENDIX**: supporting material

| additional material in support of this proposal |
| --- |
| **References:** |
| 1. **Quan PL, Junglen S, Tashmukhamedova A, Conlan S, Hutchison SK, Kurth A, Ellerbrok H, Egholm M, Briese T, Leendertz FH, Lipkin WI.** 2010. Moussa virus: A new member of the Rhabdoviridae family isolated from Culex decens mosquitoes in Cote, d'Ivoire. Virus Research **147:**17-24.  2. **Sather GE, Lewis AL, Jennings W, Bond JO, Hammon WM.** 1970. Sawgrass virus: a newly described arbovirus in Florida. American Journal of Tropical Medicine and Hygiene **19:**319-326.  3. **Wellings FM, Lewis AL, Pierce LV.** 1972. Agents encountered during arboviral ecological studies: Tampa Bay area, Florida, 1963 to 1970. American Journal of Tropical Medicine and Hygiene **21:**201-213.  4. **Ritter DG, Calisher CH, Muth DJ, Shope RE, Murphy FA, Whitfield SG.** 1978. New Minto virus: a new rhabdovirus from ticks in Alaska. Canadian Journal of Microbiology **24:**422-426.  5. **Walker PJ, Firth C, Widen SG, Blasdell KR, Guzman H, Wood TG, Paradkar PN, Holmes EC, Tesh RB, Vasilakis N.** 2015. Evolution of genome size and complexity in the *Rhabdoviridae*. PLoS Pathogens **11:**e1004664.  6. **Tesh RB, Travassos da Rosa APA, Travassos da Rosa JS.** 1983. Antigenic relationship among rhabdoviruses infecting terrestrial vertebrates. Journal of General Virology **64:**169-176.  7. **Main AJ, Carey AB.** 1980. Connecticut virus: a new Sawgrass group virus from *Ixodes dentatus* (Acari: Ixodidae). Journal of Medical Entomology **17:**473-476.  8. **Tokarz R, Sameroff S, Leon MS, Jain K, Lipkin WI.** 2014. Genome characterization of Long Island tick rhabdovirus, a new virus identified in *Amblyomma* *americanum* ticks. Virology Journal **11:**e26.  9. **Dilcher M, Faye O, Faye O, Weber F, Koch A, Sadegh C, Weidmann M, Sall AA.** 2015. Zahedan rhabdovirus, a novel virus detected in ticks from Iran. Virology Journal **12:**e183.  10. **Kerschner JH, Calisher CH, Vorndam AV, Francy DB.** 1986. Identification and characterization of Bahia Grande, Reed Ranch and Muir Springs viruses, related members of the family *Rhabdoviridae* with widespread distribution in the United States. Journal of General Virology **67:**1081-1089.  11. **Walker PJ, Kongsuwan K.** 1999. Deduced structural model for animal rhabdovirus glycoproteins. Journal of General Virology **80:**1211-1220.  12. **Roche S, Bressanelli S, Rey FA, Gaudin Y.** 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science **313:**187-191. |

|  |
| --- |
| **Annex:**  This proposal addresses the classification of several tick-borne and mosquito-borne rhabdoviruses that together form a monophyletic clade in maximum likelihood trees generated using complete L protein sequences. The viruses include Moussa virus which is currently classified to an unassigned species (*Moussa virus*) within the *Rhabdoviridae*, and 9 other viruses that are not currently classified.  **Genus *Mousrhavirus***  Moussa virus (MOUV; strain C23) was isolated in 2004 from mosquitoes (*Culex decens*) collected near the Taï National Park in Côte d’Ivoire (1). Several other MOUV isolates were obtained mosquitoes collected from the same location at that time. Bullet-shaped particles have been observed by negative-contrast electron microscopy (1). The complete MOUV genome sequence (11,526 nt) has been determined (1). MOUV is currently classified to an unassigned species (*Moussa virus*) in the *Rhabdoviridae*.  **Genus *Sawgrhavirus***  Sawgrass virus (SAWV; strain 64A-1247) was isolated from hard ticks (*Dermacentor variabilis*) taken from an opossum captured at Sawgrass Lake, Tampa Bay, Florida, in 1964 (2). Eight other isolates of the virus were obtained from the same species of tick taken from opossums, racoons and rabbits, and from hard ticks of another species (*Haemaphysalis leporispalustris*) taken from rabbits captured in the same area in 1964 (2), and further 19 isolates were obtained from hard ticks of the same two species in 1968 and 1969 (3). Bullet-shaped SAWV particles have been observed budding from neuronal plasma membranes and accumulating within extracellular spaces in the brains of experimentally infected newborn mice (4). The complete SAWV genome sequence (11,216 nt) has been determined (5).  New Minto virus (NMV) was isolated on three occasions from hard ticks (*Haemaphysalis leporispalustris*) removed from snowshoe hares (*Lepus americanus*) captured in New Minto, Alaska, in 1972 (4). The designated prototype strain of NMV (0579) was shown to cross-react weakly in complement-fixation and neutralisation tests with SAWV (4, 6). Bullet-shaped NMV particles have been observed budding from plasma membranes and accumulating within extracellular spaces. The near-complete NMV genome sequence (11,156 nt) has been determined, including complete coding sequences but incomplete 3' and 5' termini (5).  Connecticut virus (CNTV; strain Ar1152-78) was isolated from hard ticks (*Ixodes dentatus*) taken from an eastern cottontail rabbit (*Sylvilagus floridanus*) captured in Lyme, Connecticut, in 1978 (7). Neutralising antibody to CNTV has also been detected in eastern cottontail rabbits (7). CNTV cross-reacts weakly in complement-fixation and neutralisation tests with SAWV and NMV (6, 7). The near-complete genome sequence (11,169 nt) has been determined, including complete coding sequences but incomplete 3' and 5' termini (5).  Long Island tick rhabdovirus (LITRV; strain LS1) was detected in hard ticks (*Amblyomma americanum*) collected on Long Island, New York, in 2013 (8). The mammalian host was not reported. The complete LITRV genome sequence (11,176 nt) has been determined (8).  SAWV, NMV and CNTV have been described as the ‘Sawgrass serogroup’ of rhabdoviruses (7). As no isolate is available, the antigenic relationship of LITRV to the Sawgrass serogroup cannot be determined using the conventional tests.  **Genus *Zarhavirus***  Zahedan rhabdovirus (ZARV; strain Ar Teh 157764) was isolated from hard ticks (*Hyalomma anatolicum anatolicum*) collected in 2001, in Zahedan, Iran (9). The virus was shown to replicate in Vero cell cultures and newborn mice. The complete ZARV genome sequence (13,230 nt) has been determined (9).  **Genus *Lostrhavirus***  Lone Star tick rhabdovirus (LSTRV; strain TickAa42) was detected in 2009 by deep sequencing of a hard tick (*Amblyomma americanum*) collected in the USA from a patient with a rash illness. The near-complete genome sequence (11,504 nt) has been determined, including complete coding sequences but incomplete 3' and 5' termini.  **Genus *Barhavirus***  Bahia Grande virus (BGV) was isolated on 27 occasions from mosquitoes (*Aedes, Culex, Anopheles, Psorophora* spp*.*)collected in Texas, Louisiana, New Mexico and North Dakota, between 1972 and 1979 (10). The designated prototype strain (TB4-1054) was isolated from *Aedes sollicitans* mosquitoes collected at Brownsville, Texas, in 1974 (10). BGV has been reported to have rhabdovirus-like morphology in thin sections of infected cells by transmission electron microscopy (10)(8). Neutralising antibody to BGV has been detected in humans, cattle, sheep, wood rats (*Neotoma micropus*), and opossums (*Didelphis marsupialis*). The complete BGV genome sequence (12,639 nt) has been determined (5).  Harlingen virus (HARV; strain PV01-3828) was isolated from mosquitoes (*Aedes salinarius*) collected at Harlingen, Texas, in 2001. The complete HARV genome sequence (12,626 nt) has been determined (5). The HARV genome shares 92.6% nucleotide sequence identity with the BGV prototype strain (PV01-3828) and 98.3% nucleotide sequence identity with BGV strain 79V-5816, which was isolated from mosquitoes (*Aedes flavescens*) collected in North Dakota, in 1979. HARV and the prototype strain of BGV also share high levels of amino acid sequence identity in the N, G and L proteins (>97.5%) (**Tables 1-3**). HARV is therefore also considered to be a strain of BGV.  Muir Springs virus (MSV; strain 76V-23524) was isolated from mosquitoes (*Aedes* sp.) collected in Fort Morgan, Colorado, in 1976 (10). MSV has been reported to have rhabdovirus-like morphology in thin sections of infected cells by transmission electron microscopy (10). MSV cross-reacts with, but is distinguishable from BGV in both complement-fixation and plaque-reduction neutralisation tests (10). The complete MSV genome sequence (12,580 nt) has been determined (5).  Other related viruses  Reed Ranch virus (RRAV; strain TB4-222) was isolated from mosquitoes (*Aedes salinarius*) collected in Brownsville, Texas, in 1974 (10). RRAV has been reported to have rhabdovirus-like morphology in thin sections of infected cells by transmission electron microscopy (10). RRAV cross-reacts weakly in complement-fixation tests with BGV and MSV but does not cross-react in plaque reduction neutralisation tests (10). No genome sequence data are available. It is a probable member of the genus *Barhavirus*.  BGV, MSV and RRAV have been described as the ‘Bahia Grande serogroup’ of rhabdoviruses (8).  **Genome organisations**  The genomes of this set of viruses range in size and are approximately 11.2-13.2 kb in length, each containing the five canonical rhabdovirus structural protein genes (*N*, *P*, *M*, *G* and *L*). Several other alternative ORFs (>180 nt) occur variously within each of these genes (**Figure 1**).  *Genus Mousrhavirus*: There are no alternative ORFs (>180 nt) in MOUV (**Figure 1**).  Genus *Sawgrhavirus*: In most cases, alternative ORFs are unique (**Figure 1**). However, in both CNTV and LITRV, there are two alternative ORFs in the *G* gene and these share identifiable sequence homology (**Figure 2**).  Genus *Zarhavirus*: There are no alternative ORFs (>180 nt) in ZARV (**Figure 1**).  *Genus Lostrhavirus***:** Alternative ORFs (>180 nt) occur in the *N* gene and *P* gene of LSTRV.  Genus *Barhavirus*: Alternative ORFs occur in the *N* gene and in the *G* gene (overlapping the end of the G ORF) in BGV and HARV (**Figure 1**). Although BGV and HARV are considered to be strains of the same virus, it may be significant that the alternative ORFs (which share very high levels of sequence identity) are preserved in these and another sequenced isolate of BGV (79V5816) (**Figure 2**). It may also be significant that in MSV there is an equivalent alternative ORF overlapping the end of the G ORF but there is no initiation codon in the published sequence (**Figure 2**).  **Glycoprotein structures**  A Clustal X alignment indicates that the G proteins of all viruses contain all 12 cysteine residues that are present in vesicular stomatitis Indiana virus (VSIV), forming six disulphide bonds in the folded protein (11, 12) (**Figure 3**). No other cysteine residues occur in any of the ectodomains.  **Phylogenetic analysis**  Based on ML trees generated from complete L protein sequences, the viruses to be assigned to the five new genera form a well-supported monophyletic clade which is distinct from all currently assigned genera and other currently unclassified rhabdoviruses (**Figure 4**). Within this larger clade, the viruses assigned to the five genera form distinct branches or sub-clades, each with a specific host/vector association (ticks or mosquitoes). Two of the genera (*Sawgrhavirus* and *Barhavirus*) comprise viruses that previously have been assigned to distinct serogroups (Sawgrass serogroup and Bahia Grande serogroup, respectively).  **Amino acid sequence identities**  Amino acid sequence identities in pair-wise alignments (p-distances) in the N proteins, G proteins and L proteins indicate that the proposed sawgrhaviruses (SAWV, NMV, CNTV, LITRV) and barhaviruses (BGV, HARV and MSV) form well-defined groups (**Tables 1-3**).  Genus *Sawgrhavirus*: Amino acid sequence divergence is >6% in N, >20% in G and >15% in L.  Genus *Barhavirus*: Amino acid sequence divergence is >20% in N, >20% in G and >10% in L. Sequence divergence between BGV and HARV is <2% in N, <3% in G and 1% in L, confirming their assignment as strains of the same virus.  Amino acid sequence divergence between viruses to be assigned to different genera exceed 60% in the N and G proteins, and 50% in the L protein. This is in line with genus-level divergence levels for other rhabdoviruses.  **Species demarcation criteria**  Viruses assigned to different species within each genus have several of the following characteristics: A) minimum amino acid sequence divergence of 10% in N proteins; B) minimum sequence divergence of 10% in the L proteins; C) minimum amino acid sequence divergence of 15% in G proteins; D) significant differences in genome organisation as evidenced by numbers and locations of ORFs; E) can be distinguished in virus neutralisation tests; and F) occupy different ecological niches as evidenced by differences in vertebrate hosts and or arthropod vectors.  Genera *Mousrhavirus*, *Zarhavirus* and *Lostrhavirus*. At this stage, each of these genera contain only single species.  Genus *Sawgrhavirus*: All proposed members meet criterion A, except for CNTV and LITRV which diverge in the N protein by only 6.4%. All four proposed members meet demarcation criteria B and C (sequence divergence in the G and L proteins). Although there are differences in the number and location of alternative ORFs in each virus, it is not known if the ORFs are of functional significance (criterion D). Three proposed members (SAWV, CNTV and NMV meet criterion E as they are distinguishable in cross-neutralisation tests; no isolate of LITRV is available. All four proposed members appear to meet criterion F as they were isolated from hard ticks of four different genera from four distinct geographic locations in North America. However, SAWV has also been isolated on one occasion from a tick of the same species as NMV (*Haemaphysalis leporispalustris*).  Genus *Barhavirus*: The two proposed members of the new genus *Barhavirus* (BGV and MSV) meet demarcation criteria A, B and C. They also appear to be distinguishable in genome organisation (criterion D) and in cross-neutralisation tests (criterion E). They appear to share similar ecology circulating in *Aedes* mosquitoes in the USA (criterion F).  **Derivation of the genus names**  *Mousrhavirus* is derived from Moussa (a coffee plantation in Côte d’Ivoire) where Moussa virus was first isolated, and rhabdovirus.  *Sawgrhavirus* is derived from Sawgrass Lake in Florida where Sawgrass virus (assigned to the type species of the genus) was first isolated, and rhabdovirus.  *Zarhavirus* is derived from Zahedan, Iran, where Zahedan virus was first isolated, and rhabdovirus.  *Lostrhavirus* is derived from Lone Star tick (*Amblyomma americanum*) in which Lone Star tick rhabdovirus was first detected, and rhabdovirus.  *Barhavirus* is derived from Bahia Grande, a body of water near Brownsville, Texas, where Bahia Grande virus (assigned to the type species of the genus) was first isolated, and rhabdovirus.  **Type species**  *Moussa mousrhavirus* is currently the only species assigned to the genus.  *Sawgrass sawgrhavirus* is designated as the type species of the genus *Sawgrhavirus* as Sawgrass virus was the first identified of the viruses assigned to the genus. Uniquely amongst the viruses to be assigned to this genus, there are multiple isolates of SAWV from two different tick vectors. The complete genome sequence has been determined and rhabdovirus particles have been observed by transmission electron microscopy.  *Zahedan zarhavirus* is currently the only species assigned to the genus.  *Lonestar lostrhavirus* is currently the only species assigned to the genus.  *Bahia barhavirus* is designated as the type species of the genus *Barhavirus* as Bahia Grande virus was the first identified of the viruses assigned to the genus. Uniquely amongst the viruses to be assigned to this genus, there are multiple isolates of BGV from several different mosquito vectors and vertebrate hosts have been identified by serology. The complete genome sequence has been determined and rhabdovirus particles have been observed by transmission electron microscopy. |

**Figure 1.** Genome organisations ofviruses to be assigned to new genera *Sawgrhavirus*, *Mousrhavirus*, *Zarhavirus*, *Lostrhavirus*, and *Barhavirus*. Each genome contains long open reading frames (ORFs) in the *N*, *P*, *M*, *G* and *L* genes (open arrows). Other long ORFs shaded in the same colour in alternative reading frames within the *N*, *P*, *M*, *G* and *L* genes appear to encode sets of homologous proteins. Other unique ORFs >180 nt are shaded in grey but the significance of these is unknown.

Sawgrhavirus Gx proteins

CNTV MPLVGNQSWASIGFPPLFPPGGPQQSQISHAPLSEWILSLPSSSSTRPMESILRRVVPGKNTLGGCVSGKLTKQRVIQTS

LITRV ---------------------------MSH------------------MSSTPQWEVPGRSTPGGCVSGKPTKPHATRTS

:\*\* \*.\* : \*\*\*:.\* \*\*\*\*\*\*\* \*\* :. :\*\*

CNTV GVIRRLSERSGQFQLTQTSADKQSHTTSSDHMRTPSILTLHVPGWQLRIPTEQGLYCYHTRRW--------

LITRV GVTRLSSMKNGLSSLTWKNAGRQYHIISLGHTRIQDTRTRSAHGWLYRAHTEPGHCSFPTLHLLIHSVTPL

\*\* \* \* :.\* .\*\* ..\*.:\* \* \* .\* \* . \* . \*\* \* \*\* \* .: \* :

Sawgrhavirus Gy proteins

CNTV MHKCTATNTS----WD-----------------CP---------NFPHAQPLSTTSHRLLSGLQDEQWVNRIQPSTLRSH

LITRV MQRGMASNTNTCREWSQVRYRTPGADCKERMLKCPPTDKIRCFSNLPHAQSLPTPPHWILQSIPYEQWFGRVQSSAICSH

\*:: \*:\*\*. \*. \*\* \*:\*\*\*\*.\*.\*..\* :\*..: \*\*\*..\*:\*.\*:: \*\*

CNTV IQYHCPTPPRFRSSP----------

LITRV IQYYNAPQHQLRSPPQQITLPSTYD

\*\*\*: .. ::\*\*.\*

Barhavirus Nx proteins

BGV\_TB41054 MDAHHRNNAPSTKILSSQCSQKCTRPNGNNKFCNCCCHAVCCWGRVQYAICRGCTSSRSHERNANSRS

BGV\_79V5816 MDAYHRNNAPPTKILSSQCSQKCTRPNGNNKFCNCCCHAVCCWGRVQYAICWGCTSSRSHEGNANSRG

HARV MDAYHRNNAPPTKILSSQCSQKCTRPNGNNKFCNCCCHAVCCWGRVQYAICWGCTSSRSHEGNANSRS

\*\*\*:\*\*\*\*\*\*.\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\* \*\*\*\*\*\*

Barhavirus Gx proteins

BGV\_TB41054 -----------MPQHQTLSQDPPQQPIVFSTVKKVGYIILFSIPLIVIIVFFVLLDRKHLKYIPLKAWNTSIITIKEPY

BGV\_79V5816 -----------MPQHQILSQDPPQQPIVFSIVKKISYLILFSIPLIVIIIFFVLLDRKHLKYIPIKTWNTSIIEVKESH

HARV -----------MPQHQILSQDPPQQPIVFSIVKKISYLILFSIPLIVIIIFFVLLDRKHLKYIPIKTWNTSIIEVKESH

MSV LTHREPCVPFL**I**HCQNPKKKLQETCRIVFSIGKFVLKLVLFISPLIIIIVFFILLNKKHLIPISIPDINQTTVKMVPES

:.::: .\*:\* :: \*\*\*\* \* : ::\*\* \*\*\*:\*\*:\*\*:\*\*:: \*\*\* \*.: \* : : :

**Figure 2.** Clustal X alignments ofputative proteins encoded in alternative ORFs the *N* and *G* genes of sawgrahavirus and barhaviruses. In MSV, the ORF corresponding to the Gx ORF has no initiation codon in the published sequence.

**Signal domain CI**

VSIV\_G MKCLLYLAFLFIGVNC-------KFT-----------------------IVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMP-KSHKAIQA

SAWV\_G MLTHFTLLLFIPHTFG----WEPELG----------------------EYWAPIPVT-PWRPATKSDFSCPSLRVDPIPPLQLVNNTRMAYLMEGKARKEH

LITRV\_G MLRYAFVLLMIAVAFG----WEPELG----------------------EYWVPSPIS-PWRLATKTDFTCPSIRVDPVSPLKAINESYVEYPTMGGARKEH

CNTV\_G MLWSILFLVSAVYASG----WEPELG----------------------EYWVPTPVS-PWRPATKSDFTCPTLRVDPVPPLKLINETDGEYPTTGGARKEH

NMV\_G MMIVLFFFTTFTYAIGL---WEPELG----------------------VYWVPTPIT-PWKAATHNNLICPPPATRISTSLSPSFNYTVPNPRNLMGHMQH

MOUV\_G MRTLVIWVLINATMAFAKPPGSASLS--------------------LGLYWVPRIDNNTWKSVHTTNLVCPSFVGSVLPEMEESFEVDIQVPKHSQTTSHQ

ZARV\_G MVSFGLLLLFLSSASGL------NVT--------------------APFFLYPELVS-PWRNITLDHAKCPISSPVATDQELSSETFPSGHQELHHELAPV

LSTRV\_G MIRIQPGLAILIVFTISPIL---SSR--------------------HSVIMFPHVVT-PWIRAKTSDLSCPPSSPLETPNALPITEIDVFIQKDHKELELI

BGV\_G MISNMFFLFQLSLFLQFIAG---DES-LETITAPETPDPILLKGDTKYLFLVPSSVK-NWKPADLNELTCPPLISKPDTSEMTYFSTDVMELQKHHELAPV

HARV\_G MISNMFFLFQLSLFLQLIAG---DES-LETITAPETPDPILLKGDTKYLFLVPSSVK-NWKPADLNELTCPPLISKPDTSEMTYFSTDVMELQKHHELAPV

MSV\_G MATFNIIFVLISFWTTLGIS---DESPHITVTAPETPDPILLQGDKTYLFLVPSESK-NWKPADLNELSCPPLISKPDTAEMEYMSTDVMELQKHHELAPV

. \* . \* \*\*

**CII CIII CIV CV CVI**

VSIV\_G DGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTGEWVDSQFIDGKCSN

SAWV\_G PGWLCVRKMYQTTCDTNFWGHQTITHEELALPADQTECRQAIAHYIKGTYEDPRHPDPVCTWMAAEKVYNPNTALLPHSTVVDPFSYTFMDSLFPGTHCKT

LITRV\_G PGWLCIRKTYQTTCDTNFWGHQTIKHEEWSIVADVEECRQAVSHYQLGTYEDPRHPDPVCTWMAISSTYRAGTLLLPHTTLVDPFSYTFVDSLFPGTHCKT

CNTV\_G PGWLCVRKTYQTTCDTNFWGHQTIKREEWPISADPDECRQAISHYQLGSYEDPKHPDPTCTWMAIAHTYRAGTLLLPHSTVVDPFSYTFMDSLFPGIHCKV

NMV\_G DGWLCVSSLYATTCDTNFWGHQTITQSNLPVRMTSAACKKAVREHILGELAPPRYPDPYCYWMASHTQAITQIRVIEHPVTSDLYTETFVNSLFPGTHCAM

MOUV\_G GGYLCYGFSFSVVCEEGFWGGQKVTEHTFTHLVSSEECLKAIEDKKSGEYRPPHTPVSECGWMQTNTKTLRFVALEEHPVLFDPYTVNFVDGLFEKTLCNQ

ZARV\_G KGYLCSGLRYTTECYEGFLGGKDIKKTIDKVQVTGDICLRELEKVKTGSVIPPYFPAPECAWMKTQTNDAVFYIIEEHSVSYDPYKIGFVDPIFLKDICVN

LSTRV\_G PGLLCYGLKYVTHCSEGFFGQKTISKRIDKIAPTLTDCRKSYEIYEKGEQIDPYFPASKCTWMAEDEDSKNFYLLTPHSVKYDPYTTGATDPLFLRDHCKE

BGV\_G EGYLCSGLRYKVICSEGFFGQKTIAKKIENIEPDSKQCLDDLSKFKNDDYLLPYFPSEDCNWMKETPTHKDFIVFQKHFVKYDPYNNGFYDPLLKKDYCDT

HARV\_G EGYLCSGLRYKVICSEGFFGQKTITKKIENIEPDSKQCIDDLSKFKNDDYLLPYFPSEDCNWMKETPTHKDFIVFQKHLVKYDPYNNGFYDPLLKKDYCDT

MSV\_G QGYLCSGLRYKVICSEGFFGQKTITKKIENLEPDQNKCVQDLEKFINDDYLLPYFPSEDCNWMKETPVHQDFIVYQKHQVKYDPYHNGFYDALFKKDFCQE

\* :\* : . \* : \* : : . \* . \* \* \* : \* . \* : : : \*

**CVII CVIII CIX CX CXI**

VSIV\_G DICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGDKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEG

SAWV\_G APCPTVHPDVLWHTVTPLTEDCPGHTIIPIKIYNDKSARKKTH----EWLSVSDGPLIPLGGSCSLTYCGQLGLRTPGGHWYP-----YPGSKRYPECKDA

LITRV\_G IPCMTVHPDVLWHSSSTITENCPMANGIHIKLYNDNRARKKAH----EWLSVNDGPLIPLGGSCSMHYCGQAGLRTPGGIWYP-----YLGRQKYPECKEA

CNTV\_G LPCATVHPDILWHSESTIVEDCPMTQAIHLKLYNDKSARKKAH----EWLSVHDGPLIPLGGSCALSYCGRSGVRTPGGVWYP-----YPGNRVYPECKDA

NMV\_G NPCTTVHPDTMWETTTIIKKDCKILPNSNFTGYKD--PRRPDR----EWVVIDNDSAIYLAGSCHMQYCGQKGLRTASGQWIP-----YEGPTKYPDCPES

MOUV\_G RICPTVHANTIWIGDNEPKKDCPSTENEKAVLYVEK-QNVVPV----VWVKLTGGTVYKLDRACTMTYCDIDGVRMEDGHWFAGVN--LTQYVR-RNCDKG

ZARV\_G KVCPTEHDNMLWVTEADVGHQCSEFLDAEITFHSN---QNYSI----TIIEHQHKMFEDFTKACRMTYCGEKGVRLPSGLFYTHIP--PPYWHNINECSVK

LSTRV\_G TTCETVHANTIWMTTKSIN-RCSPFQRHVGTLYHD---PRKNV----THIQYKGTHTYSLNGSCLMNYCDKFGVRLSNGLFLGGPL--ATKIPNGPKCPQG

BGV\_G QVCETEHDQTIWITEKSIENECIFNYPIKKHIFHT---ADFGK----MIIDYELNQWTSVEDGCLINYCGREGIRLSNGMFFVGKF--YKNLNNLQTCSAG

HARV\_G QVCETEHDQTIWITEKSIENECIFNYPVKKHIFHT---ADFGK----MIIDYELNQWTSVEDGCLINYCGREGIRLSNGMFFVGKF--YKTLNYLQTCSAG

MSV\_G KICETEHDQTIWITNQELKQECTFNYPVKKHVFYK---RDYSK----MIIDYEINQWTSVEDGCLIRYCGQEGIRLSNGMFFVGKF--YKLISNLPICPEG

\* \* \* . \* \* . . .\* : \*\* \*:\* .\* : \*

**CXII**

VSIV\_G SSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVEIAAPILSRMVGMISGTTTERELW

SAWV\_G WRATPTFE-GEGMRLDIELQEILARRECINVLQQIQSGAAPTFHMLSHFQPRHAGYYRAYRMKNNVIEYSLALYDPIFNVSVPPQVNFGAWQN-GTRYYLP

LITRV\_G WRATPTPA-GNGAKLDIELQELIARRECLNVLQQIRSGVSPTFHMLSHFQPRHTGYYRVYRMNNGLVEYSLALYAPIFNITMPPSINFGVRRN-KSRYHLP

CNTV\_G WRATSTPA-VDGVKIDIELQEIMARRECINVLQQIRAGIAPTFHMLSHFQPRHTGYYRVYRMNNGLIEYSLALYAPIFNITVPPRLDFGVRRN-KSHYHLP

NMV\_G WLTRSTEH-SNELKVNLEVQEVESRRACIEATQRIRDGAPISFHLLSYFQPRRTGYYHVYRIYKGILQYSKAWYEPLKDLNPSGKYTLGHFPN-SSIYKID

MOUV\_G MDITFDTLASLSLLTKIELEHVQDRMECLDAVQDLRAGGKVTYAKLSKLQPRRGGLFHVYRINKGTLEYTMGRYEGLTSLITNIPFVIGKNQK-DEKVRLP

ZARV\_G RNVTFYPLSAELLEIEREMTVDRERMMCIESLQTARRTKVMTFLTLSYLVPKFEGRFPVYRLEKGQLKGAVANWHALKTVKPGTSRQIGTFPNGTNAYWWD

LSTRV\_G TTVKFVPIEDEIETLEEEIQEDRDRETCLVQLMTVRMSNFSTFFSLSYMDPKFESRGKVFRIHQNHLEMAHADWIPVTNVTNHKQNLIGISNRGKELYFTD

BGV\_G TKVSYKPLTSKLEEIENEIILDQERLLCLDSIRQMTATKKLSFYSLSFLEPKSSSRHKVFRIHNKTLEYTETEWHPIMSFNFDEPNKIGIDKNGKSVYWNE

HARV\_G TKVSYKPLTSKLEEIENEIILDQERLLCLDSIRQMTATRKLSFYSLSFLEPKSSSRHKVFRIHNNTLEYTETEWHPIMSFNFDEPNKIGIDKNGKSVYWNE

MSV\_G TKISYKPIKAQLDEIENEIILNQERLLCLDSIRQMTASKKLSFYSLSFLEPKSMSRHKVYRIHNNTLEYTETEWEPIVAFNFNGKNQIGVNKEGKEVYWNE

:: \* : \*\* : \*: . .: : : :: : : .\*

**Transmembrane**

VSIV\_G DDWAPYEDVEIGPNGVLRTS--SGYKFPLYMIGHGMLDSDLRLSSKAQVFEHPHIQDAASQLPDDETLFFGDTGLSKNPIELVEGWFSGWKSSIASFFFII

SAWV\_G LVESGYPGTLSAFNGIHVMH-NKTVVVPELELYKEHYAETLFYYTAQIAEHPSQMRQDNETELTPPTRTTITVKRPSIGGWLSSMWGSVWGKVTSAIGLVL

LITRV\_G MIESGHPGVWSAFNGIHIMH-NKTVIVPELELYKEHYSETLFYYKAQLAEHPSQVRQANSTDITPHTKTTVTVKRPTLKAWFSTMWDSLWGKVVSITGIIL

CNTV\_G LIKSGHPGVWSAFNGIHVMH-NKTVIVPELEFYKEHYSETLLYYRAQVVEHPTQVRQANSTELTPHKKTSLIVNRPTLGNWLSLIWDSFWGKLVSILGIVS

NMV\_G PVRADKNGTLSAFNGVHVAP-DGTIIVPEVELFKDTYSDTLLYQKARLIDHPATAIQANYTTLVPHYTSTTSFHRPDLTAWGASVWSAFWGKVMLISMAVA

MOUV\_G HIPSGDNSTLSSYNGVHMFL-NGTVIIPEMELYKLRYSETLLYEHLLGEMKHPSAKQRERMGLTPDDDKRTTNKSLNIGEWFTSFWSHLVGKIVSILGTAL

ZARV\_G WVDSGYPGIDSGFNGVHRT--EGKVVIPRLEVLKKEYALSLDILHEMRTIEHPIIKHLARENLTGHLTTLEGRDSIDVGEWLSNLWEKIWGKLLLLGCLCG

LSTRV\_G WVPSGRHNLSSGWNGVHKTS-KGKIIIPRLSLLKQEYEETLLVEHNLREITPVHIHHFDKEGLNDTVVTLTSHELVDVGAWIGSVWSTLWGKITTIVTVII

BGV\_G WVPSGISGLLSGFNGVYKKENETKVTIARLETIKEDYDREMMIDHELVEVEHPKIVHLKRENITGSRVEIVNKEHSDVSGWLSSVLSSFWGKIMMTIISII

HARV\_G WVPSGIPGLLSGFNGVYKKENETKVTIARLETIKEDYDREMMIDHELVEVEHPKIVHLKRENITGSRVEIVNKEHSDVGGWLSSVLSSFWGKIMMTIISII

MSV\_G WVPSGKDGLLSGFNGVYKKVNSSKISISRLETIKEDYEREMMIDHELVTVEHPKIVHLKRENITGSRVEIVNTEHSDVSGWFSSVLKSFWGKLMMTVVSII

: . . \*\*: .. : : . .

**domain**

VSIV\_G GLIIGLFLVLRVSIYLCIKLKHT-------------------------------------------KKRQIYTDIEMNRLGK------------------

SAWV\_G AIFVIGLLIKILP---WGRLLRR-------------------------------------------PSTPAPRTVVYTPATGNVAW--------------

LITRV\_G AIFTTGFLIKVLP---WSRLFRR-------------------------------------------PQPMHPQIVHYTPATGSVSW--------------

CNTV\_G AVVITGILIRLLP---WGRLLRK-------------------------------------------PHPTRPKIVHYTPATNNVSW--------------

NMV\_G SVLIIYVAIKCVP---WAAITRR-------------------------------------------RAQPMPAVVTYTPSNSRVNW--------------

MOUV\_G AIFLILYICWICLKIQIKRVSDKNR-----------------------------------VDQMEMQILSKARAPEVRPTLSGPIW--------------

ZARV\_G ---FLLVVCCCCFGRIRRCLSTTSRQNND---------------------------------IEMLPLR---TNLETVKIDRPVHW--------------

LSTRV\_G ALIILYFIVRCCVPILSRCCKKKESKHPE---------------------------------FEMVPLRSPSSTPSRVTLTRPKGWQ-------------

BGV\_G LIVIIGLVLINCCPIICKSCIKRYKTKEESRNRHRLDREDNGRLRRQHRVIFNNQSNDEENAIEMVEYTDTPRPLRPIPDATTSDTESRSPTTAHSFFNR

HARV\_G SIIIIGLVLINCCPIICKSCIKRYKTRKESRNRHRLDREDNGRLRRQHRVIFNNQANDEENAIEMVEYTDTPRPLRPIPDAPTSDIESRSPTTAHSFFNR

MSV\_G IIIIIGLLIINCGPIICKTCISSYKKKKSRRDRFRADRETETGLRRQHRVVFHNNETDDERAIEMMEYSDTPRTLRPIPDSLPEPQEETTRNMSHSFFNR

**Figure 3.** A Clustal X alignment of the G proteins of VSIV and viruses to be assigned to new genera *Sawgrhavirus*, *Mousrhavirus*, *Zarhavirus*, *Lostrhavirus*, and *Barhavirus*. Conserved cysteine residues in the VSIV G protein are marked (CI-CXII).



**Figure 4.** The evolutionary history was inferred from a Clustal W alignment of 137 complete L protein sequences of 128 animal rhabdoviruses currently assigned or recently proposed for assignment to species and proposed members of the new genera *Sawgrhavirus*, *Mousrhavirus*, *Zarhavirus*, *Lostrhavirus*, and *Barhavirus*. Phylogenetically informative sites were selected from the alignment using Gblocks resulting in 964 positions in the final dataset. The tree was inferred in MEGA7 by using the Maximum Likelihood method based on the Whelan and Goldman + Freq. model. The tree with the highest log likelihood (-109033.5365) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrap values (100 iterations) are shown for each node.

**Table 1.** Percentage amino acid sequence identities (p-distance) of a CLUSTAL W alignment of the N proteins ofviruses to be assigned to new genera *Sawgrhavirus*, *Mousrhavirus*, *Zarhavirus*, *Lostrhavirus*, and *Barhavirus*.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SAWV | LITRV | CNTV | NMV | MOUV | ZARV | LSTRV | BGV | HARV | MSV |
| SAWV |  |  |  |  |  |  |  |  |  |  |
| LITRV | 89.5 |  |  |  |  |  |  |  |  |  |
| CNTV | 88.4 | 93.6 |  |  |  |  |  |  |  |  |
| NMV | 68.7 | 69.6 | 69.4 |  |  |  |  |  |  |  |
| MOUV | 38.4 | 37.9 | 36.8 | 35.8 |  |  |  |  |  |  |
| ZARV | 25.8 | 24.9 | 24.9 | 24.4 | 27.9 |  |  |  |  |  |
| LSTRV | 21.9 | 21.2 | 21.0 | 21.7 | 26.6 | 32.9 |  |  |  |  |
| BGV | 23.3 | 23.5 | 23.3 | 21.9 | 23.5 | 29.5 | 32.0 |  |  |  |
| HARV | 23.3 | 23.7 | 23.5 | 21.7 | 23.7 | 30.1 | 32.2 | 98.9 |  |  |
| MSV | 24.0 | 23.5 | 24.4 | 23.3 | 25.3 | 30.8 | 32.6 | 79.7 | 79.9 |  |

**Table 2.** Percentage amino acid sequence identities (p-distance) of a CLUSTAL W alignment of the G proteins ofviruses to be assigned to new genera *Sawgrhavirus*, *Mousrhavirus*, *Zarhavirus*, *Lostrhavirus*, and *Barhavirus*.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SAWV | LITRV | CNTV | NMV | MOUV | ZARV | LSTRV | BGV | HARV | MSV |
| SAWV |  |  |  |  |  |  |  |  |  |  |
| LITRV | 68.7 |  |  |  |  |  |  |  |  |  |
| CNTV | 67.4 | 76.1 |  |  |  |  |  |  |  |  |
| NMV | 29.0 | 40.0 | 39.0 |  |  |  |  |  |  |  |
| MOUV | 26.8 | 27.6 | 27.2 | 29.0 |  |  |  |  |  |  |
| ZARV | 24.3 | 23.4 | 25.1 | 23.2 | 25.5 |  |  |  |  |  |
| LSTRV | 25.9 | 25.9 | 26.8 | 26.3 | 27.2 | 35.5 |  |  |  |  |
| BGV | 22.2 | 21.8 | 23.2 | 21.4 | 24.9 | 35.1 | 34.4 |  |  |  |
| HARV | 22.6 | 21.8 | 23.4 | 21.6 | 25.1 | 34.6 | 34.6 | 97.7 |  |  |
| MSV | 23.3 | 21.6 | 21.8 | 23.2 | 25.1 | 33.4 | 33.6 | 78.2 | 78.8 |  |

**Table 3.** Percentage amino acid sequence identities (p-distance) of a CLUSTAL W alignment of the L proteins ofviruses to be assigned to new genera *Sawgrhavirus*, *Mousrhavirus*, *Zarhavirus*, *Lostrhavirus*, and *Barhavirus*.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SAWV | LITRV | CNTV | NMV | MOUV | ZARV | LSTRV | BGV | HARV | MSV |
| SAWV |  |  |  |  |  |  |  |  |  |  |
| LITRV | 79.8 |  |  |  |  |  |  |  |  |  |
| CNTV | 80.0 | 85.0 |  |  |  |  |  |  |  |  |
| NMV | 69.3 | 70.5 | 70.5 |  |  |  |  |  |  |  |
| MOUV | 51.6 | 52.1 | 52.4 | 52.0 |  |  |  |  |  |  |
| ZARV | 45.0 | 45.0 | 45.2 | 45.3 | 45.5 |  |  |  |  |  |
| LSTRV | 45.0 | 45.2 | 44.7 | 44.3 | 47.7 | 55.5 |  |  |  |  |
| BGV | 43.6 | 43.7 | 43.0 | 43.7 | 45.7 | 52.9 | 53.7 |  |  |  |
| HARV | 43.4 | 43.6 | 42.9 | 43.7 | 45.6 | 53.0 | 53.6 | 99.0 |  |  |
| MSV | 43.1 | 44.1 | 43.1 | 44.1 | 46.2 | 53.8 | 53.3 | 85.2 | 85.5 |  |