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:** | ***2019.004M*** | | | | (to be completed by ICTV officers) |
| **Short title:** Create one new genus (*Sunrhavirus*), including six new 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  Juliana Freitas-Astúa  Hideki Kondo  Gael Kurath  Ivan V. Kuzmin  David M. Stone  Robert B. Tesh  Noel Tordo  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:** | | | | | |
| Supported by 12 of 13 SG members with limited corrections related to formatting and spelling errors. There was one non-responder. | | | | | |
|  | | | | | |
| Date first submitted to ICTV: | | | |  | |
| 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:** 2019.004M.A.v1.Sunrhavirus.xlsx |

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

| additional material in support of this proposal |
| --- |
| **References:** |
| 1. **Ledermann JP, Zeidner N, Borland EM, Mutebi JP, Lanciotti RS, Miller BR, Lutwama JJ, Tendo JM, Andama V, Powers AM.** 2014. Sunguru virus: a novel virus in the family *Rhabdoviridae* isolated from a chicken in north-western Uganda. Journal of General Virology **95:**1436-1443.  2. **Karabatsos N (ed).** 1985. International Catalogue of Arboviruses Including Certain other Viruses of Vertebrates. American Society for Tropical Medicine and Hygiene, San Antonio.  3. **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.  4. **McAllister J, Gauci PJ, Mitchell IR, Boyle DB, Bulach DM, Weir RP, Melville LF, Davis SS, Gubala AJ.** 2014. Genomic characterisation of Almpiwar virus, Harrison Dam virus and Walkabout Creek virus; three novel rhabdoviruses from northern Australia. Virology Reports **3:**1-17.  5. **Humphrey-Smith I, Cybinski DH, Byrne KA, St George TD.** 1991. Seroepidemiology of arboviruses among seabirds and island residents of the Great Barrier Reef and Coral Sea. Epidemiology and Infection **107:**435-440.  6. **De Haas RA, Jonkers AH, Heinemann DW.** 1966. Kwatta virus, a new agent isolated from *Culex* mosquitoes in Surinam. American Journal of Tropical Medicine and Hygiene **15:**954-957.  7. **Karabatsos N, Lipman MB, Garrison MS, Mongillo CA.** 1973. The morphology, morphogenesis, and serological characterization of the rhabdoviruses Navarro, Kwatta, and Mossuril. Journal of General Virology **21:**429-433.  8. **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.  9. **Calisher CH, Karabatsos N, Zeller H, Digoutte J-P, Tesh RB, Shope RE, Travassos da Rosa APA, St. George TD.** 1989. Antigenic relationships among rhabdoviruses from vertebrates and hematophagous arthropods. Intervirology **30:**241-257.  10. **Bourhy H, Gubala A, Weir RP, Boyle DB.** 2008. Animal rhabdoviruses, p 121. *In* Mahy BWJ, Van Regenmortel MHV (ed), Encyclopedia of Virology, Third Edition, vol 1. Elsevier Academic Press, Oxford.  11. **Roche S, Rey FA, Gaudin Y, Bressanelli S.** 2007. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science **315:**843-848.  12. **Walker PJ, Kongsuwan K.** 1999. Deduced structural model for animal rhabdovirus glycoproteins. Journal of General Virology **80:**1211-1220. |

|  |
| --- |
| **Annex:**  The new genus *Sunrhavirus* is proposed to accommodate six currently unassigned rhabdoviruses that have been isolated from mosquitoes and birds. Each virus will be assigned to a new species within the new genus.  Sunguru virus (SUNV) was isolated from a domestic chicken in Arua District, Uganda in 2011(1). It was shown to replicate in vertebrate cell lines (BHK, Vero, chicken embryo fibroblasts) but poorly if at all in several arthropod cell lines. Infection without dissemination was observed in mosquitoes (*Anopheles gambiae*) (1). The complete SUNV genome sequence (11,056 nt) has been determined  (1).  Garba virus (GARV) was isolated from a malachite kingfisher (*Corythornis cristata*) trapped at Bangui in the Central African Republic in 1970 (2). It was subsequently isolated from a beautiful sunbird (*Nectarinia pulchella*), also from the Central African Republic (2). It has been reported to cross-react in complement-fixation and neutralization tests with Matariya virus and Burg el Arab virus, each isolated from lesser white throats [*Sylvia curruca*] in Egypt, in 1961/2 (2). The near-complete GARV genome sequence (10,821 nt), including the complete coding sequence, has been determined (3).  Harrison Dam virus (HARDV) was isolated from *Culex annulirostris* mosquitoes collected at Beatrice Hill in the Northern Territory, Australia, in 1975 (4). Trace levels of neutralizing antibodies to HARDV have been detected in horses and crocodiles (4). The near-complete HARDV genome sequence (11,284 nt), including the complete coding sequence, has been determined (4).  Walkabout Creek virus (WACV) was isolated from biting midges (*Culicoides austropalpalis*) collected near Samford in southern Queensland, Australia, in 1981 (4). All 138 sera collected from sea birds from the Great Barrier Reef and Coral Sea tested negative for WACV neutralizing antibodies (5). The complete WACV genome sequence (11,214 nt) has been determined (4).  Kwatta virus (KWAV) was isolated from *Culex* spp. mosquitoes collected near Paramaribo, Surinam, in 1964 (6). Bullet-shaped KWAV virions have been observed in ultrathin sections of infected mouse brain and infected Vero cells (7). KWAV has been reported to cross-react strongly in complement-fixation tests and neutralization tests with an unnamed and unclassified rhabdovirus (BeAn 157575) which was isolated from a white-shouldered fire-eye (*Pyriglena leucoptera*) in northern Brazil (8, 9). The near-complete KWAV genome sequence (11,211 nt), including the complete coding sequence, has been determined.  Oak Vale virus (OVV) was first isolated from *Culex edswardi* mosquitoes at Peachester, Queensland, Australia in 1981/82. The virus was subsequently isolated from *Anopheles annulipes* mosquitoes collected in 1993 near Kunnanurra in northern Western Australia. There is a report of neutralizing antibodies to OVV in feral pigs (10). The complete OVV genome sequence (11, 220 nt) has been determined (4).  The genomes range in length from approximately 10.8 kb to 11.2 kb, containing the five canonical rhabdovirus structural protein genes (*N*, *P*, *M*, *G* and *L*) (**Figure 1**). Each of the viruses also features an additional gene (*U1*) between the *M* and *G* genes containing a single ORF encoding a small hydrophobic (SH) protein. The KWAV and OVV SH proteins share identifiable sequence homology (**Figure 3**). The SH proteins of SUNV, GARV, HARD and WACV are also homologous but, although they are similar in predicted structure, they do not display obvious sequence homology with the KWAV and OVV SH proteins (**Figure 2**). Interestingly, the SUNV, GARV, HARD and WACV SH proteins also display identifiable sequence homology with the SH proteins of tupaviruses (**Figure 3**) which are also encoded in a gene (*U1*) located between the *M* gene and G gene (3). However, the tupaviruses appear to be quite distinct phylogenetically from the sunrhaviruses (**Figure 3**).  Uniquely for HARDV genome, there is also an additional gene (*U2*) between the *G* gene and *L* gene. Although this appears to be a complete and independent transcriptional unit (i.e., bounded by conserved transcription initiation and transcription termination/polyadenylation sequences) the gene is very small, encoding a protein of only 19 amino acids (2.38 kDa) (**Figure 2**).  Alternative ORFs (>180 nt) occur in the *N*, *P* and *G* genes of some sunrhaviruses but it is not known if these are expressed (**Figure 2**). Notably, there are alternative ORFs (Px) commencing near the start of the *P* genes of all sunrhaviruses encoding proteins. Although there are significant size variations in the proteins, pairwise alignments indicate that there is obvious sequence identity between the Px proteins of [HARDV and WACV], [SUNV and GARV], and [KWAV and OVV]. There was no clear indication of global sequence identity between the proteins. There is also identifiable homology between proteins encoded in ORFs located near the 5’ end of the HARDV, WACV and OVV G ORFs (alignment not shown).  Alignment of sunrhavirus G proteins with the G protein of vesicular stomatitis Indiana virus (VSIV) indicated that eight of 12 cysteine residues in the ectodomain are fully conserved, and these correspond to four established disulphide bridges (CI-CXII; CIII-CV; CVI-CVII; and CIX-CXI) (11, 12) (**Figure 4**). Sunrhaviruses all lack two cysteine residues that form another known disulphide bridge in VSIV (CVIII-CX). Cysteines comprising the sixth disulphide bridge in VSIV (CVIII-CX) are absent from KWAV and OVV, and the bridge appears to be rearranged in SUNV, GARV, HARDV and WACV (**Figure 4**). All sunrhaviruses have an additional two cysteine residues in the ectocomain which are likely to form an additional disulphide bridge (**Figure 4**).  Based on well-supported ML trees generated from complete L protein sequences, sunrhaviruses form a monophyletic clade that is distinct from all currently recognized genera and other currently unassigned rhabdoviruses (**Figure 3**). Amino acid sequence divergence in pair-wise alignments (p-distances) are >10% in the N and L proteins and >20% in the G proteins (**Tables 1-3**).  **Other possible members of the genus**  Based on antigenic cross-reactions and/or phylogenetic analysis, using partial sequences, as well as host/vector associations, several other viruses, are considered to be possible members of the new genus. These include several viruses isolated in the Central African Republic: Sandjimba virus (SJAV), isolated from a sedge warbler (*Acrocephalus schoenobaenus*) in 1970; Kolongo virus (KOLV) and Bimbo virus (BBOV), each isolated from yellow-crowned bishops (*Euplectes afra*) in 1970; Ouango virus (OUAV) isolated from a black-headed weaver (*Sitagra melanocephala*) in 1970; Boteke virus (BTKV), isolated from mosquitoes (*Mansonia maculipennis*) in 1968; and Nasuole virus (NASV), isolated from a little greenbul (*Andropapus virens*) in 1973. Matariya virus and Burg el Arab virus (BUAV), each isolated from birds collected in Egypt (see above) are also possible members of the genus. These viruses do not yet qualify for classification as only limited sequence data are currently available.  **Species demarcation criteria**  Viruses assigned to different species within the genus *Sunrhavirus* 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 organization 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.  All proposed members of the new genus meet demarcation criteria A, B, C, D and F. Cross-neutralisation tests amongst the viruses proposed to be assigned to the genus have not yet been reported (criterion E).  **Derivation of the genus name.**  *Sunrhavirus* is derived from Sunguru virus, which is assigned to the type species of the genus, and rhabdovirus.  **Type species.**  *Sunguru sunrhavirus* is designated as the type species of the genus as Sunguru virus is arguably the best characterized of the viruses currently proposed to be assigned to the new genus. An isolate is available, the complete genome sequence has been determined and experiments have been conducted to characterize growth of the virus in cell cultures and in insects. |



**Figure 1.** Sunrhavirus genome organisations. Open arrows indicate the locations of long open reading frames (ORFs), each of which is located within a transcriptional unit bounded by conserved transcription initiation and transcription termination/polyadenylation sequences. N, P, M, G and L represent ORFs encoding the canonical rhabdovirus structural protein genes. The U1 ORFs (shaded blue) also occur in independent transcriptional units and encode small hydrophobic proteins of unknown function. The HARDV U2 ORF (shaded orange) also occurs as an independent transcriptional unit and encodes a very small protein (2.38 kDa) of unknown function. Alternative ORFs (shaded purple, green and brown) occur near the start of each of the P genes. Alternative ORFs (shaded grey) of significant length (>180 nucleotides) also occur in the N, P and G genes but the significance of these is unknown.

***Sunrhavirus***

KWAV\_U1 MAGWKLLFVLLIVLYWHNPEGVTSLMKSSLNIMETILAEPIRKVVSFFTPP--CPPCPQCLVKTP

OVV\_U1 M-FWKVFFALVLFSYWNNPDVAS---RTATTIFDMMLLA-TRYIASYILPASACPPCPEVHAPFP

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

GARV\_U1 MIALFLLLTLLVMVLRPRYVEWILFYMLGHYNV-GLNALYNVNFLFWYLFCDIPSRFLNNVFGDMIEKYYQD--------

SUNV\_U1 MI-LLVVLILFFGMLYKRLSFLMMAYLLGYYNVFGSVITYG-SFFIWYLFYYLPSKWMGAGFAAIVESYNKEYAEWVSIE

HARDV\_U1 MI-IIIFLILISLFLSKRIGALIAMYTLGYYNAFGNIINYL-TFFGWYIGVNLPYKYLSQHWSELVSEYQNSKRE-----

WACV\_U1 MI-VVIIIVLLALIISKRLSAFIAMYTLGYYNAFGSLINYL-TFFGWYVGVNLPYKYLSQHWSELVAEYRNSQ-------

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

***Tupavirus***

TUPV\_U1 MI-TTLIIIGAAFLVGPRTFKFVLAYLLGYYNAFGPPLQIV-QFMVWLIIIYFPKKFFSLGWYFCHDAFSSYFGDPNGGQLPVSTKFHSLTDMID

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

**Figure 2.** ClustalX multiple sequence alignment of the small hydrophobic (SH) proteins encoded in the *U1* genes of sunrhaviruses and tupaia rhabdovirus (TUPV; *Tupaia tupavirus*).

HARDV\_Px MSQIQVTEILETLNQCHSWNCQEVSHGNHLLNWWNSCKDLTWLMVERVMSLIRGLQTRFIMMLLSTQRGRNAAI------

WACV\_Px MNPTQVRTVLETLNQCHNWNCQEESHGNHLLNWWNSCKNLTALMLERVMSLIHGLQTRFIMILLSTQRGKNAAIELTKMI

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

HARDV\_Px ----------------------

WACV\_Px VERTLTPAAEPESSTSKDMKME

GARV\_Px MSHHSVETIVR---------ILEECT---------DLECQEHPQEVHFFNLFRWGKFMIQELLRKVKSLCHGIQMKILTL

SUNV\_Px MSQQQIECVAFRRPSPVNNAMMDQTISWLDQIQRTSESWEDRSTNPPIWTLWQLVKQVFHGIGSKVQCFFHGIQIRLLET

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

GARV\_Px ILSTRPGIRAAAQTMRTVMEKQQNLVPTLQEVATETDSETECEKP------

SUNV\_Px ILSTREGRYQAIEVVKRIANQHPELFNQIEIPPLEREMETIYNSNEGGTLQ

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

KWAV\_Px MMTRKRLASGSIDRILQRKRCNHQSHQWTPQVSGWIHRQVG----KVTQAIHGVLCYLLEKILMTPVGIQVVCQILEGEE

OVRV\_Px M-TKTVLQRRSLRAALSAKRCSHQTH-LTPRVKSWITSQVERVMDRLTRVLNGIILFLLEQVIRTTPGQEAVCQLLDQVE

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

KWAV\_Px VDEDDLPQSLRMESHRPIQVLEPENPRISPVEILGQDPKQVKEAQSPPLPPPRRSPTPPDPSPIPEREDHRRMDPHSQ

OVRV\_Px LEDQ----------------VDSSEEEVD-------------------------------------------------

:::: ::..: .:.

**Figure 3.** ClustalX multiple sequence alignment of sunrhavirus Px proteins, encoded in the first alternative ORF of each of the *P* genes.

**CI**

VSIV\_G MKC---------LLYLAFLFIGVNCKFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQAD

HARDV\_G MKK------MFFIFIVISFVSYINCVEHVFFPIEMKSQFKPVKIEDLSCPYNQFERDHNNDLKIDVEILKNN---LITKT

WACV\_G MNNFCNMKKTFILIFAFACLYTAHSLEHVFFPVEMKSQFKPVKIEDLSCPYNQFERDHTSDLKIDVELLKNN---LITKT

SUNV\_G MSL-----LGLAVVSFISILSNVGSVEHLYFPVEMKSQFKPVKLEDLTCPYASDDMGFPSSVKADVQLLKTN---LIRVP

GARV\_G MKL---LIIEFFAIITSFFASRCLSLEYMYFPTSVTKGFTPIHLDHLNCPYDIDDTEIENPVEVDCKILVTN---LIDVE

KWAV\_G MDK-------LIILTACLLGVVIASHDYYYFPVVQSKSFKKLPVGQLRCPPHSSEKPLSHKKIWGGYVLTQN---IQTMP

OVV\_G MAA-------KLLIVISLCGLVLG--DYIYYPVTIQTPFKATPLSNLQCPRHPSERSLKNSGYGRGWILKLN---IVDVP

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

**CII CIII CIV CV**

VSIV\_G GWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVL

HARDV\_G GNLCYKQKWITRCEENFFGIQTLNHSIIDLPLDD----HPTSSDSN-----PLFPPPDCRWLSSSEVSRDYIICKQETIR

WACV\_G GNLCYKQRWSTKCEENFFGIQKLNHSIIDLPLDE----HPSSSDSN-----PLFPPPDCRWMSSSEVSKDYTICKQETIR

SUNV\_G GTICYRQLWTIKCSENFFGVQTINKDIKDLELGH----FPEKGDDK-----VYFPDPTCRWMSESETTAEFAICKDEEIL

GARV\_G GEVCYKQKWVTNCYENFIGQQTINHRIHHLPISADDIVHKLSLDSN-----LVPPDANCQWMSDTETEDTKVICQPVTIK

KWAV\_G GTFVVKQRWGTTCTMNFWGVKTIRHHIIDEQILD---ARFTNITLK-----PVFPDEDCSWMTTATREITYYVGTKGELE

OVV\_G GLFIVKQKWKTHCFMNFLGIKTIRHEIISETLTH---DDVKTFTTT-----PTFPNEECHWMSDTVTEKTYFIASRGTMA

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

**CVI CVII CVIII**

VSIV\_G VDEYTGEWVDSQFIDGKCSNDICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETG

HARDV\_G FDEVLGIGVDEQYGTFKCGEKFCR-PDKYITFLPIGGVDKMKLEGFEKVKGYLTLDQ---DGFVTPSSLIYSNHFPKMSL

WACV\_G FDETLGIGVDEQYGTFKCEKSYCR-LDKYITFLPIGGVPQMRLEGFEKVKSYITLDQ---EGFVTPSSLIYSNHFPKMSL

SUNV\_G FDETTGLGTDNQYGSFFCQKDYCP-INKYVGFKPKRPLAEIIKEGFMDIEAEFSVNS---RGFVDIHSLTRSHHYPRMSM

GARV\_G YDETLNLGSHPSLGTFPCVKPPCS-IDKQHVFNSTDFTATN--KGYKDTKVQFSTDQ---NGHIYETSFVKSDVFPKMSL

KWAV\_G YDISTGKTSDPVFGAFSCTEKLCY-VDHRVVFIP-DVAIAATSKGFKFVVFEISTDP---DGVIRENSVIQSRDFPRMSL

OVV\_G YDISTGRSADPVYGTHLCSMHVCH-LRSDVIFKS-DEEFKIKEYGFKEVVYEAELDE---NKQVTESSIVQSRDFHPLSI

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

**CIX CX**

VSIV\_G DKACKMQYCKHWGVRLPSGVWFEMAD------------------------------------------------------

HARDV\_G EKACLRWVEKET-YKYDVELIMNNGFLLRLPLNLVFKERGKPDTPFKPVVETNMIGHKYQIHDVLELLLARAKGTNQETK

WACV\_G KESCMRWVDKGD-YNYDIELITNNGFLLRLPLNINLK--AKSGAILTPTIENNHISHTYHINDVIELLLARAKGIQKERK

SUNV\_G KNACVRWRDHE--SKKKFDLILNNGFLIRFKS--------EFDIEVSPRLESNWVTKDYNTDDEVKKLIN-AKSHPNHGT

GARV\_G EGACIEMTKSASNNNHAAKIILQSGLLLEVRDAFDLG-------SDDSYATGNSIHVNHNLAELKKLLLKKSKKNSTWGQ

KWAV\_G RKACVT--EESVLGQRRLAFILRNGFFLVLEMGVKSG---------SHMLKKSTETLGSELILRASLRLS---NDKFKG-

OVV\_G RGACID--EQNLEG-KKFSIIFPNGFYLVLDIPIVSG---------GLNIKYE---QGQDLQKRASMRLKGTGSGPLRGV

:\* . .

**CXI CXII**

VSIV\_G ----------------------------KDLFAAARFPECPEGS--SISAPSQTSVDVSLIQDVERILDYSLCQETWSKI

HARDV\_G PFGFTRNNKWLKLDNGQKFDDYFGKSK-PLAHLLTMIRDCEEKDVQKITIPTMDFQTIEAEMFVESKIDQMFCKHRLYDI

WACV\_G LFTFTRNHRWLKLADNSKLEDKFAKKDTPFIHFFTMIRDCEEKDEKRITVPTMDFQTIEAEMFVESKIDQMFCKHRLYDI

SUNV\_G HFSFQRD-PWAKISDS-----VIKKEKFKFAHLILSLRNCEKEDDKRIKIPYIDFQTTDVEMYIESKIDQIACKRRLYEI

GARV\_G VMFDQNSRSKMTFSSD------VTGKFKQFGKLFVDLRICQKNDFKRVVVPTLDFQRSMTEMFVESKIDQLSCKKRLYEI

KWAV\_G RDLSMLY-TQEKISGAG-----------SIDNLLNGFRVCDASDRSRIKQVGLGFNSLEQDERIMSRVDSLFCRVTLDRI

OVV\_G KDVGVIFNTKVTDSKCGMWCS-----------AINGLPVCETFARHHIRQTGMLSMEFQQERRMLAKVESFICREKLHDV

: : : : :: \*: :

VSIV\_G RAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVEIAAPILSR---MVGMISGTTTERELWDDWAPYED---

HARDV\_G MTKKRANLNDLALLSPNHGGLGPVYHRYAKDITRGIGLYRRINWQPSEKG----LGYYFQNGSQIWVKCPEWVNGSE--G

WACV\_G MTKGKASLNDLALLSPNHGGLGPVYHRYANDITRGIGVYRRINWHPSKEG----LGYYFKDSKQVWVKCPEWVNGTD--G

SUNV\_G ITEKKFNLIDLGLLAPNHGGLGPVYHSYSKDISRGFGHYRRIIWDPQPEIG--ILGYYFDGGEQKNVTCPEWVRKNH--S

GARV\_G MSTGSITGADLGLLSQNHEGPGPVYQLYKHDVTMAYGTYERFTWMPKTEGGKQYLGYVHIGNTKKWVECPEWVPDSESTD

KWAV\_G RKCKKLTSVELGMFAQNYGGPGPVYRIKNDTLEVAQGIYKRIFWDPDTKNR---LGYYVNETTEKEVNCPEWIKISE--G

OVV\_G RNRKRINPISLGMFVQRHGGPGPVYRIKNKTLESATGIYKRVFWEP-TDNR---IGKYYNGTTEKEIVCKEWIQDEN--G

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

VSIV\_G VEIGPNGVLRTSSGYKFPLYMIGHGMLDSDLRLSSKAQVFEHPHIQDAASQLPDDETLFFGDTGLSKNPIELVEG-----

HARDV\_G FRWCVNGIFELNGKVYHPYFGADNFQELIKSFEENEIRKVEHNVILHDMNNRIHETWESYLGTKEDYHWKGINL------

WACV\_G FRWCVNGIFELNGKVYHPYFGADNFQELIKSFEEDEVRRVKHNVILHDLGSRSHETWESYLGAKENYTWKGFNL------

SUNV\_G LSWCVNGIFKMGGKIFHPIYGADNLEELKIAFEERDVRSVEHPAILHDLHNREATTWKEYHKVTDDLRWKGVNLGIWDFF

GARV\_G IKWCVNGIFERNGSLYHPVFGGDNIRDLKVAYQVRNLRKVEHLSLLLQTNNRTISNWEEYFKLESKHTFEGWNR-----V

KWAV\_G FESCINGIIRYKNVTSHPLSPVNDLEQEEALFKEHFLEDVYHVPTQHLN---PWAGWNPLHPPEIDRHFLGLKLP-----

OVV\_G -HSCVNGIVRYGKKIIHPSSLANTAPEEERLFAESELVDAYHVPTKAVN---PWADWNPLHPPPSHRKFLGLHFP-----

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

VSIV\_G --WFSGWKSSIASFFFIIGLIIGLFLVLRVSIYLCIKLKHTKKRQIYTDIEMNRLGK

HARDV\_G -GWIHSIQEWAKNIWIIIIVILTVFALLMLIKLTR-----GKRRNRHDW--------

WACV\_G -SWIDSIRKWANNSWMIVVGVCSGLGILLLIKITR-----GKRSGGADW--------

SUNV\_G DSVIGKIVACALGGLMFLILFWVSASVIMRCCVKRSFHNPNAKRRSEEW--------

GARV\_G SHWFKGLSDEIKYIFYGLSFVFVLFCIVRLLRWRH------RSQLYLDY--------

KWAV\_G -----NIFGFMHNFEIYLVTFIVGLISLPLIIFCC-------RRKSSRY--------

OVV\_G -----DFLGSVKYYLEWILIGSLGFLLMLLIISCR-------NRNRGYY--------

: :

**Figure 4.** ClustalX multiple sequence alignment of the G proteins of sunrhaviruses and vesicular stomatitis Indiana virus (VSIV). Twelve cysteine residues in VSIV form six disulphide bridges (CI-CXII; CII-CIV; CIII-CV; CVI-CVII; CVIII-CX and CIX-CXI) and these bridges are conserved to various extents amongst rhabdoviruses in patterns that are somewhat genus-specific (11, 12). The figure illustrates conservation in the sunrhaviruses of the cysteine residues forming four of these bridges (CI-CXII; CIII-CV; CVI-CVII; and CIX-CXI). Cysteines forming the CVIII-CX bridge are absent in all sunrhaviruses and the CII-CIV bridge is disrupted. Cysteines CII and CVI are absent in KWAV and OVV; although cysteine CII is present in HARDV, WACV, SUNV and GARV, the CVI cysteine is absent and appears to be replaced by a cysteine residue that follows CV. All sunrhaviruses also have two additional conserved cysteines downstream of CXII that likely form an additional disulphide bridge. Unusually, GARV and KWAV have single unlinked cysteine residues in the ectodomain. OVV has two additional closely spaced cysteine residues that may form a unique bridge. Predicted N-terminal signal domains (dark grey) and C-terminal transmembrane domains (light grey) are also shown.



**Figure 3.** The evolutionary history was inferred from a Clustal W alignment of complete L protein sequences of 6 proposed sunrhaviruses and 113 other animal rhabdoviruses currently assigned or recently proposed for assignment to species. Phylogenetically informative sites were selected from the alignment using Gblocks resulting in 1053 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 (-109153.9075) 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-Join 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 sunrhavirus N proteins.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | SUNV | GARV | HARDV | WACV | KWAV | OVV |
| SUNV |  |  |  |  |  |  |
| GARV | 39.5 |  |  |  |  |  |
| HARDV | 47.7 | 37.6 |  |  |  |  |
| WACV | 49.6 | 37.8 | 86.7 |  |  |  |
| KWAV | 33.5 | 28.9 | 29.2 | 29.9 |  |  |
| OVV | 31.1 | 28.2 | 30.1 | 30.8 | 38.6 |  |

**Table 2.** Percentage amino acid sequence identities (p-distance) of a CLUSTAL W alignment of sunrhavirus G proteins.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | SUNV | GARV | HARDV | WACV | KWAV | OVV |
| SUNV |  |  |  |  |  |  |
| GARV | 34.6 |  |  |  |  |  |
| HARDV | 45.6 | 36.7 |  |  |  |  |
| WACV | 46.0 | 35.9 | 76.9 |  |  |  |
| KWAV | 26.8 | 26.0 | 28.2 | 25.6 |  |  |
| OVV | 24.3 | 24.1 | 25.6 | 24.1 | 42.3 |  |

**Table 3.** Percentage amino acid sequence identities (p-distance) of a CLUSTAL W alignment of sunrhavirus L proteins.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | SUNV | GARV | HARDV | WACV | KWAV | OVV |
| SUNV |  |  |  |  |  |  |
| GARV | 51.4 |  |  |  |  |  |
| HARDV | 55.2 | 54.5 |  |  |  |  |
| WACV | 54.2 | 54.6 | 85.7 |  |  |  |
| KWAV | 40.6 | 41.3 | 42.3 | 41.4 |  |  |
| OVV | 39.4 | 40.4 | 40.8 | 40.5 | 55.9 |  |