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:** | ***2019.006M*** | | | | (to be completed by ICTV officers) |
| **Short title:** One new genus (*Arurhavirus*), including 4 new species, in the family *Rhabdoviridae*. | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| Members of the ICTV *Rhabdoviridae* Study Group:  Peter J. Walker  Kim R. Blasdell  Ralf G. Dietzgen  Juliana Freitas-Astúa  Hideki Kondo  Gael Kurath  Ivan Kuzmin  Robert B. Tesh  Nikos Vasilakis | | | | | |
| **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:** | | | | | |
| Proposal supported by a majority of Study Group members (9 supporters and 4 non-responders). | | | | | |
|  | | | | | |
| 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:** Arurhavirus\_Excel\_module.xlxs |

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

| additional material in support of this proposal |
| --- |
| **References:** |
| 1. **Spence L, Anderson CR, Aitken TH, Downs WG.** 1966. Aruac virus, a new agent isolated from Trinidadian mosquitoes. American Journal of Tropical Medicine and Hygiene **15:**231-234.  2. **Spence L, Jonkers AH, Grant LS.** 1968. Arboviruses in the Caribbean Islands. Progress in Medical Virology **10:**415-486.  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. **Aitken TH, Woodall JP, De Andrade AH, Bensabath G, Shope RE.** 1975. Pacui virus, phlebotomine flies, and small mammals in Brazil: an epidemiological study. American Journal of Tropical Medicine and Hygiene **24:**358-368.  5. **Karabatsos N (ed).** 1985. International Catalogue of Arboviruses Including Certain other Viruses of Vertebrates. American Society for Tropical Medicine and Hygiene, San Antonio.  6. **Wanzeller AL, Martins LC, Diniz Junior JA, de Almeida Medeiros DB, Cardoso JF, da Silva DE, de Oliveira LF, de Vasconcelos JM, Nunes MR, Vianez Junior JL, Vasconcelos PF.** 2014. Xiburema Virus, a hitherto undescribed virus within the family *Rhabdoviridae* isolated in the Brazilian Amazon region. Genome Announcements **2:**e00454-00414.  7. **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:**  The new genus *Arurhavirus* is proposed to accommodate four currently unassigned rhabdoviruses that have been detected in mosquitoes and sandflies from the Caribbean and South America. Each virus will be assigned to a new species within the new genus.  Aruac virus (ARUV) was first isolated from mosquitoes (*Trichoprosopon theobaldi*) collected from Melaju Forest, Trinidad, in 1955 (1). The virus was also subsequently isolated from 1955 to 1963 from mosquitoes of several other species (*Wyeomyia* sp., *Psorophora ferox*, *Phoniomyia* sp., *Culex* sp., *Sabethes chloropterus*) collected from the same region of Trinidad (1). There is a report of low prevalence neutralising antibodies to ARUV in birds (species not reported) (2). The near-complete genome sequence (11,906 nt) of the virus has been determined, including complete coding sequences but incomplete 3' and 5' termini (3).  Inhangapi virus (INHV) was isolated from sandflies (*Lutzomyia flaviscutellata*) collected in the Catu Forest, Belem, Para State, Brazil, in 1969 (4). Neutralising antibodies to INHV have been detected in rodents (*Proechimys guyannensis*, *Hylaeamys megacephalus*, *Coendou* sp.) (5). The complete genome sequence (12,026 nt) of the virus has been determined (3).  Xiburema virus (XIBV) was isolated from mosquitoes (*Sabethes intermedius*) in Sena Madureira, Acre State, Brazil, in 1974. The virus has been reported to have rhabdovirus morphology by electron microscopy (5). The near-complete genome sequence (12,240 nt) has been determined, including complete coding sequences but an incomplete 3' terminus (6).  Santa Barbara virus (SBAV) was detected in sandflies (Psychodidae) collected in Santa Barbara do Para, Para State, Brazil, in 2010 (Genbank entry only). The near-complete genome sequence (12,162 nt) has been determined, including complete coding sequences but incomplete 3' and 5' termini.  Arurhavirus genomes range in length from approximately 11.9 kb to 12.3 kb, all containing the five canonical rhabdovirus structural protein genes (*N*, *P*, *M*, *G* and *L*) and an additional gene encoding a viroporin-like protein (**Figure 1, Figure 2**). In ARUV, XIBV and INHV, the viroporin-like protein is encoded in an additional transcriptional unit (gene) (*U1*) between the *G* gene and *L* gene; In SBAB, the viroporin-like protein is encoded in an alternative ORF (Gx) within the G gene overlapping the end of the G ORF. In the ARUV *U1* gene, encoding the 14.1 kDa viroporin-like protein, contains a second ORF (U1x) of 156 nt encoding a 5.9 kDa protein of unknown function. The U1 termination codon and the U1x initiation codon overlap and are preceded by a TURBS (termination upstream ribosome-binding site)-like sequence (AGGGA) suggesting that expression of ORF U1x occurs by a stop-start mechanism as observed for similar arrangements is several other rhabdoviruses, including ephemeroviruses, hapaviruses, curioviruses and sripuviruses. The XIBV *U1* gene, encoding the 11.4 kDa viroporin-like protein, is followed by a gene (*U2*) that contains an ORF of 65 nt encoding a 7.8 kDa protein of unknown function. There is no significant sequence homology between the ARUV U1x and XIBV U2 proteins. The SBAV genome, uniquely amongst the arurhaviruses, contains an additional long ORF of 243 nucleotides in an independent gene (*U1*) between the *N* and *P* genes, encoding a 9.0 kDa protein is of unknown function. Other ORFs (>180 nt) in alternative reading frames occur variously in the *P*, *G* and *L* genes of the arurhaviruses but the significance of these is unknown. Note that the convention for accessory gene nomenclature described in Walker *et al*. (3) has been adopted here.  A Clustal X alignment indicates that arurhavirus G proteins contain all 12 cysteine residues that in vesicular stomatitis Indiana virus (VSIV) form six disulphide bonds in the folded protein (7) (**Figure 3**). SBAV has two additional cysteine residues (**e** and **f** in **Figure 3**) that are likely to form an additional disulphide bond in the ‘stalk’ region of the glycoprotein.  Based on ML trees generated from complete L protein sequences, arurhaviruses form a well-supported monophyletic clade that is distinct from all currently assigned genera and other currently unassigned rhabdoviruses (**Figure 4**). Amino acid sequence divergence in pair-wise alignments (p-distances) are >30% in the N proteins, >60% in the G proteins and >30% in the L protein (**Tables 1-3**).  **Other related viruses:**  An important consideration in the formation of this new genus is the relationship of the proposed arurhaviruses to those already assigned to the genus *Curiovirus*, i.e., Curionopolis virus (CURV; species *Curionopolis curiovirus*), Iriri virus (IRIRV; species *Iriri curiovirus*); Rochambeau virus (RBUV; species *Rochambeau curiovirus*) and Itacaiunas virus (ITAV; species *Itacaiunas curiovirus*). The arurhaviruses and curioviruses form adjacent clades in ML phylogenetic trees inferred from complete L protein sequences, although the connecting node is relatively deeply rooted (**Figure 4**).  *Similarities.* Arurhaviruses and curioviruses appear to have similar aspects of ecology. The curioviruses were isolated from biting midges (CURV, ITAV) and sandflies (IRIRV) collected in Brazil, and from mosquitoes collected in French Guiana (RBUV). Neutralising antibody to CURV has been detected in birds. Arurhavirus and curiovirus genomes are of similar size and have some common features. Members of each genus contain the major structural protein genes (*N*, *P*, *M*, *G* and *L*) and an additional gene located between the *G* and *L* genes encoding a viroporin-like protein. [However, it is important to note that viroporin genes do occur in this location in many other rhabdoviruses, e.g., almendraviruses, ephemeroviruses, tibroviruses, hapaviruses.] As in ARUV (but not other arurhaviruses), the gene encoding the viroporin-like protein in several curioviruses (CURV, IRIRV, RBUV) features a second overlapping ORF that appears to be expressed by a TURBS-like mechanism. However, the corresponding ARUV protein (U1x) is much smaller and displays no identifiable homology with the curiovirus proteins.  *Differences.* Unlike the arurhaviruses, curioviruses have the common characteristic of one or more ORFs in additional genes located between the *M* and *G* genes (**Figure 1**). The proteins encoded in first of these ORFs (U1) are homologous in all curioviruses. CURV, IRIRV and RBUV all share common proteins encoded in the U1x and U2 ORFs. In arurhaviruses, there are no ORFs between the *M* and *G* genes (**Figure 1**). The curioviruses also display significant differences from arurhaviruses in the size and structure of their G proteins (**Figure 3**). In each genus, the G proteins contain each of the 12 conserved cysteine residues (CI-CXII) that form 6 disulphide bridges supporting the folded secondary structure of the VSIV G protein (7). However, the curioviruses each have the unusual feature (amongst animal rhabdoviruses) of a greatly extended N-terminal domain upstream of CI that is also relatively rich in cysteine residues. This domain is of unknown function and does not occur in arurhavirus G proteins.  On balance, the significant differences in genome organisation and G protein structure, indicate that arurhaviruses should be recognized as a new genus separate from curioviruses.  **Species demarcation criteria**  Viruses assigned to different species within the genus *Arurhavirus* 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.  All proposed members of the new genus meet demarcation criteria A, B, C, and D. Cross-neutralisation tests have not been reported (criterion E). All except INHV and SBAV appear to occupy different ecological niches based upon their location and source (host/vector) of isolation/detection (criterion F). INHV and SBAV were each detected in sandflies near Belem in Para State, Brazil.  **Derivation of the genus name.**  *Arurhavirus* is derived from Aruac, an ancient native tribe of Trinidadian Americans after which Aruac virus (assigned to the type species of the genus) was named, and rhabdovirus.  **Type species.**  *Aruac arurhavirus* is designated as the type species of the genus as Aruac virus was the first identified of the viruses assigned to the genus. Uniquely amongst the viruses to me assigned to this genus, there are multiple isolates of ARUV from multiple insect vectors and the susceptibility to infection of a range of laboratory animals has been evaluated. |



**Figure 1.** Arurhavirus and curiovirus genome organisations. Each genome contains long open reading frames (ORFs) in the *N*, *P*, *M*, *G* and *L* genes (open arrows) and an ORF encoding a viroporin-like protein (shaded yellow). Other long ORFs are shaded to show sets of homologous proteins in the same colour and ORFs encoding unique proteins in other colours. ORFs >180 nt in alternative reading frames within the *P*, *G* and *L* genes are shaded in grey but the significance of these is unknown.

ARUV\_U1 MHNQTRAVVSFPNLPNVGAFADSIKNTLEETWIRFVQKIESWSDTVKQTLIVVFIIICIFIAIKI

XIBV\_U1 MERGIGFNPSQAIDGIKGALNNLGNSISSFFNDIGIKLNYWGKIFLIIIGVILGLIIIPRLISNV

INHV\_U1 MNITSRINWQSVDPSKWFEGIKDGANSFFSSLKVVFQDTKYWINLFFWLIVAILIMVIVLKFSNY

SBAV\_Gx MERGVFDPLVNWLSQASNNVKFYFWVIWLLMLTLGFLFVLSKAIDLISSIILYATQFYKAFQISL

ARUV\_U1 LIVCGQLCTTICQGTT**H**GLNCI**R**SCS**RRR**IL**R**PSPPP**K**P**KRRK**QTSAPD**R**YYGYAVTQV

XIBV\_U1 CTIISGLS**K**CL**R**WLW**R**MVSAC**R**L**R**CP**R**LCC**K**F**RR**GDN

INHV\_U1 LIGLINQCGACV**K**MIGNCQC**K**C**KKK**A**KKR**G**K**II**K**IN**K**IYNP**R**NM

SBAV\_GX **R**WIT**KK**C**K**NC**K**TNNNQISN**KR**GL**R**SAVALV

**Figure 2.** Arurhavirus viroporin-like proteins. Like similar proteins in other rhabdoviruses, these proteins have the characteristics of class 1a viroporins, featuring large hydrophobic residues (F, W, Y) in the N-terminal domain, a central hydrophobic transmembrane domain (shaded light grey) and a C-terminal domain that is rich in basic residues (K, R).

**Signal peptide domain**

VSIV\_G MKCLLYLAFLFIGVNC-----------------------------------------------------------------

**XIBV\_G** MGICLPGHHPPIPTTKKPSV-------------------------------------------------------------

**ARUV\_G** MTMRQWVLLVLFLLGVFP---------------------------------------------------------------

**INHV\_G** MSKSIVIFLICFVFITQVSS-------------------------------------------------------------

**SBAV\_G** MKLFVLAIVCLGLRSRVINC-------------------------------------------------------------

CURV\_G MDLVRFSIALSVFLCYGTPPSQG---QAIVSIKDSCEAKSAPWIPCEKFDYVKNATGSGIKCWIFCSRSGFYSKTGRFIRC

RBUV\_G MKKTMELIKVIPVACLLYIFLS---------------------------------------------------------RH

IRIRV\_G MKKTMETIKNFILILSVAVQLDTGHGMEFALIRNTCDTRSMKQVPCQHLISVRESTGPKGKCLIFCAKEKRRWAEGVESQC

ITAV\_G MKDLHTLVLLVVLIKLYKGGSS------VHTPMTRCTVYAPQQRIQCNRFDKVFRIRGGGVGIILCRRS------------

\*

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

**XIBV\_G** ----------------------------------------------------RNIARLP--REILNR---PEMKDIGVVII

**ARUV\_G** ---------------------------------------GNVVVLRHDQELNPEYGRKMKVTEHTFRGKDGIWRKFLVTIF

**INHV\_G** -------------------------------------------ILDWPFRF-PEIGKMIAQNDVLKKGTDGKATKYRLSLI

**SBAV\_G** ----------------------------------------------------QDKQRYVITNKGWLY---DTVDNWYLEPY

CURV\_G IQGDPEAKYIKSCRRQIEKRGKEKMREGTRGKRKTSEPKEEGVRAKTDFTP-DESRRLNNLTKVFRKVEDKDLNDFKKFIL

RBUV\_G IL-------VGQCKDKIQTG-KNTSPPKSR-NFKGSPPDSDSVLAKTNLTQ-EESKRLSNISQVLEKIEDRDLAEFKRIIL

IRIRV\_G LQGDPTSPQVVKCKERIEKGLERLTVLTSSKRKRMREP----FRARGNQTK-EEGGRLIRISDVLKRIDHKDLEQFKKDVL

ITAV\_G ---QKFDPNLCTCFRDTNPGALSCYKLMTSKGSFPSEKESRGDEKPQESGTLADYLNVTFLEKTGKLGAKKGEDPDESRLL

**a b**

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

**XIBV\_G** DRRLDGSSIHG------------------------KVERATIQRKQQNPQEVR-------------------EEE------

**ARUV\_G** GKQVFGSDYPER-----------HSIAESFVETTVTPEHSRQKRATPDPNAMYS--------------KLKRAGG------

**INHV\_G** AIPLGTFGRPP----------YDIAEPG----TPMPVTLPRVTRDVKPLEHLI------------------ESQG----QY

**SBAV\_G** -RPFGPP----------------------------KTIERQIEEKLNLTKGNIP------------------NKG------

CURV\_G EKGLETKIKLANDGKISFRD-PDCGENKDYPCHRIHQIIEGVNENIDYINEIL---SLKKMKEELRLRERESEEGEFPGLL

RBUV\_G DKNLEKSLSEDKDGKIHLNA-PKCGENDSYRCHRVNKILENVNDNIDHLNDLLRKSSGSNLREKLRELDDSSEGG---GKG

IRIRV\_G EKNVETFLYVNSSNLIDMHKNYSCSPPDGFECYSIKRAIDDSNEKVLYVQKAQKEKDKSTWSKDLLSDAYRREEG----TI

ITAV\_G DAAIKSGGLGVDFNGHLKILNEESLTSNLNETDVNRIKGLLTSWNKDRLGGQPGEYDPNKWKKDAKEMTEEDENS----ER

**CI CII CIII**

VSIV\_G -------KFTIVFPHNQKGNWKNV-PSNYHYCPSSSDLNWHNDLIG----TALQVKMPK-SHKAIQADGWMCHASKWVTTC

**XIBV\_G** YDQVEGKVTYMNWP-SKSGEWTENFGPTGPVCPRFSSQFYITHDKL----IRAVLTRPE-HPEGVFVDGYLCEAQIWSSKC

**ARUV\_G** YQDSPGWESYVSWP-KKLGEWTSNFAGGGPKCPIFKDPLSVHSRSE----ILIHMQRPS-HPEGIFVQGYLCEAQSWMSRC

**INHV\_G** IQTDRSVRYYLNWP-DQIQAWSNKFPDDGPICPQFKTSLKVIDRKL----TKMTMSRPM-HPEGLFVNGYLCESQVWISRC

**SBAV\_G** SYPSESKMYMYNWP-FGPYLWHELTPDSNIGCPKFKSNQGELNNIN----TKITVKRPS-HPDGLFVKGYICEGLEWVLTC

CURV\_G NTTNRRG-FLLHYP-VELGNWSRLEDPSQIKCPSHHKDMLSNPRRLGKYNLDIIVRRPRIGTFETVVPGYICQGMQWTSTC

RBUV\_G SKSDGHGSFLLHYP-ITVGNWSELEDPTLIKCPSHHRNQIGHSGSDRRLDLDIIARRPRTSTFQTSVPGYICHGMRWTSSC

IRIRV\_G ETATERG-FLIHYP-IESRDWSNLTDPSQIKCPSHQGDRLRKSGTLRNVDLEIIVRRPRPGNFKTAVPGYICQGMQWTSTC

ITAV\_G HRVNSTRLRMIFHP-EGPYNWTAVTDPQRIRCPRPRLYSKSVDHNFRG-----VAKRPR-VTETSLVRGYICEAHRWVSEC

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

**CIV CV**

VSIV\_G DFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTGEWVDSQFI

**XIBV\_G** TKTWYFSYQEERSISGAVAEVEKCFQALELQKAGNPVKPEHPLSVCYWNAEHTESVVYHLITPHTSTVNPYTNKIVDPLFL

**ARUV\_G** KETWYFSYEETQKVEKQTAELEKCHEAIRLLEEGNPVIPRYPLALCYWNADHTETIVYHTVTRHTAALNPYTSHLVDPVML

**INHV\_G** KETWYFSYDEKREIKYDVPVYEKCLESLRLYQKGDPIELEHPLMVCYWNAEHVESRVYHVITPHVTTINPYRNEVQDPLLA

**SBAV\_G** SETWYWSVSFTKQIKSFTPTLDECQYSVKAYESGEPITPYFPEPLCFWNADHTESIKFTITTQHSTLFDPYRGMLDDHTMV

CURV\_G NEMWYFVTYHDRAVHYITPNKLKCLQNIRAHKRGEHIKPYYPLEECNWNSETTKTVDYFMITPYSPEVDPFTLEFKSEIFP

RBUV\_G NEMWYFVTYHDRAVTSITPNKLKCIQNIQAWKRGEHVKPYYPLEECNWNAETTKTVDYHMIIPYSLRVDPLSLQFEGEIFL

IRIRV\_G NEMWYFVTYHDRAVKTITPNEVKCLQNIQSHKRGEDVRPYYPLEECNWNAETTKTVDYHMITPYSPLVDPFTLEFKSEIFL

ITAV\_G EEMWYFSKYYTREIQTTVPEVLACIKAIKDLKSGEPEIPTFPLSLCNWNSKTSQGVIFHTATPVDILLDPFSMTYKSTLFP

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

**CVI CVII CVIII**

VSIV\_G DG-KCS--NDICPTVHNSTTWHSD----YKVKGLCDSN------LISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGDK

**XIBV\_G**  KG-GCSLDSHICQTSKLSVMWLRTGHD--SPSMVCNVDNWKKQQIVLHEVDMET---HEKGKWARRTILEGDHLGTKVLES

**ARUV\_G**  RG-NCALNSTACPTSHASVLWIKDSHH--SDHTVCTKEDWEDYEMSVHLTGLQV---AGN-KWETVHLLEGEGIGTKVLEK

**INHV\_G**  NR-TCQLKNSHCSTIRESTIWIRDSHD--PLEGICNLKNWNHAEVDVSEVDSET---SIQYRWRKGHHLEGPEFGMKLLES

**SBAV\_G**  VN-TCPYNSTYCKLKDVNKIWVRDDNDKYRTHEICKIDTWEEFDANVYFLDLPDNANSLDANWKNHLLIEGEGIGKKDLSR

CURV\_G DRTSCRPGDEICVTDDDSKVWFPDEDDKLIARGHCPDETWDESHLTIHPEEMPENWEDPQSPWVSDYILKGVLFGEKRVKK

RBUV\_G DSQSCKPGDSLCFTDDNSKIWFPDDDDKLVATGHCPDESWDESSLTIHPEKMPNDWSNENTSWSQDYILKGILFGEKRVRR

IRIRV\_G ESSSCSPGDKFCLTDDTSRIWFPDQDDPLIALGHCQDTTWDESFLTVHPEEIPRNWSDPNSTWTNDLILKGILFGEKRVSK

ITAV\_G EKWPCNAHALYCLTSKPWRKWFPDEELTRLSPQLCKTTSWEAFPFFGDVIDLPSKPRDPSSLWTPHVLVENEIFGQKNLND

\* \* . ... :. :

**CIX CX CXI CXII**

VSIV\_G ACKMQYCKHWGVRLPSGVWFEMADKD-----------LFAAARFPECPE---GSSISAPSQTSVDVSLIQDVERILDYSLC

**XIBV\_G** GCRTKFCNLNGIRFDDSEWWTMDPHEDFQHNPLYGRIRDMISQLPICSEEK-SKDIGIAHPNFETSFLKKEMGNMLRALRC

**ARUV\_G** SCKMCFCGTTGIRFDDSEWWTFKGYTQSQKGGAEIRFKDVMEGLPDCDQRK-SKEIGISHPHFSESYGRKEVGNMLRALQC

**INHV\_G** GCQMTFCGIRGVRFTDGEWWTSLVTDPTRNQTKSNQIYNAISDLKTCSAEE-SSDIGIAHPNFQDPVRKKEIGNVVKALQC

**SBAV\_G** ACRMFFCGIPGIRFPDREWWSIKIDGHNNEVEYQN-----LDMIPECTN---NQTIGLSHPHFKDSEEKIEVLDIIKTVRC

CURV\_G SCLLEFCGTSGLLFEDGEWWELNVFSREKGRESLTKIFIEQEEIRRCNGTE-TR-VGVAGKETDEKALLNAVLSKNAYERC

RBUV\_G SCLLEFCGAMGLLFEDGEWWEISVYSADSKRESLTKLFLKEEGINRCNGTE-ER-IGVAGKETDEKALLDAVLRKTGYEKC

IRIRV\_G SCILEFCGELGLLFEDGEWWEINVFSKEQHRSRLTKQFLEEEKLERCTGTE-SR-IGVVGKETDEKALLDAALRRSGYEKC

ITAV\_G GCLMDFCGETGVKFPDGEWWLFSVLHGPRGG-----LKNLTSALSRCKNSSSLHQIAEVFPNLEDSGRRLELTSRERYERC

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

VSIV\_G QETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTL---KYFETRYIRVEIAAPILSR--------------------

**XIBV\_G** NDAVGKILEGESITPVDLSALAPTSPGPGIMYKLEETEG---GQIVLKWAHANYSLIRVSP--------------------

**ARUV\_G** NNAIGKLMEGEHLTPIDISSLAQEKPGLGLMYRLHRDKS---GFFDLQWAHANYSLILYKP--------------------

**INHV\_G** SQTIGKLLSGEELTPLDLSSLAPDIPGPGIVYKLKWVND---SGYILHWAHATYKLIRYHPRQI-----------------

**SBAV\_G** AETLSKIYASFQISPFDLSALAQEHPGIGLVFSLRKQDN--GRSPKIYWSYANYKLIEVSLDQNSEFDTSVLGLKQNRTNL

CURV\_G KSARYRLIENKYLRLDDLSYINPRESVTWWAYRVRAGDD--ERTFKLEKTTGEYRYLQVPPSLE-----------------

RBUV\_G KAARYRLTENKHLRLDDLSYINPRDSVMWPAYRVRRQSEGSEKYFSLETAMSEYGMVQVTRSLQ-----------------

IRIRV\_G KAARYRLIENRNLRLDDLSYINPRDSVNWWSYRVRAEDS--DRLFRLEKAISLYRLVQVTPKLL-----------------

ITAV\_G LDVKDKLTAGYFMNPTDLGYIGQHQEGPGTAYRLKRLGTNKTGQLILMESPAVYRVLIHKEDYDEE---------------

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

**c d e f**

VSIV\_G -----------MVGMIS--------------GTTTERELWDDWAPYEDVEIGPNGVLRTSSG-------------------

**XIBV\_G** -----------IPNTDELGYAKG--------GSKIIIKEWTNTTIPGIQ-VGPNGILKKHLM-------------------

**ARUV\_G** -----------DPLLNSIGLDYL--------QNPVTLDEWVTSDIPGIQ-IGVNGIIRRIDK-------------------

**INHV\_G** -----------HPGNISIGYDHL--------NKEMLVNEWENTPVPSIK-IGHNGVLKKIVTP------------------

**SBAV\_G** FVDSRLNEHQNETMPITTSPKREGSLIGWYHNQEVWSNEWYAGSVAGLHEVGLNGIVRESLGDSKQCVKNGATLNCSNKPS

CURV\_G -----------QHVTDCDGQENCSVSIGYYRGELINSSDWTRTGHDDVY-VGVNGLLRKDTG-------------------

RBUV\_G -----------MLGQKCYGVENCSMIVGYNLGKEIRSDDWTPSGHPGVY-VGINGLIRKDDE-------------------

IRIRV\_G -----------KGGQVCEDKRNCSLIIGYHKGKEIKSSDWEETGHDGVY-IGVNGLIRRDTR-------------------

ITAV\_G -----GFRVWENSNSSWNESGKLSVQFGITSKGLKFEAEWTSLKSEGRW-VGVNGILLLKGEK------------------

\* :\* \*\*::

VSIV\_G ------YKFPLYMIGHGMLDSDLRLSSKAQVFEHPHIQDAASQLPDDETLFFGDTG---LSKNPIELVEGWFSGWKSS---

**XIBV\_G** -NGSSVLLAPQSMLQLGELDPSFSRRIDVNHIKWTEDTGAETLSEEELIPLKNGS---IIRHNAMDWIEMWVHEVGDVVSG

**ARUV\_G** -NKTLEILAPQIMLQMGELSPEFAKKIAVKRIDWIDDKGESNIIDDDGVDIPRGE---LDRTNVVDGVRNWVEKVRDTISG

**INHV\_G** -GNNFSILVPQMMLQMGEVDNSFVNKIPVEHIKLEQINDEPIIDDDEKNMIKNDEKDVINRENVVDGVGSWLHKIGEVITG

**SBAV\_G** SITKYKIIFPSAAYKMGLVTSMITKGIPISNILRPTDMTQKEFDLENPNDIVRETDS-INRVDIVNSTKHWWNGITHSFSF

CURV\_G --NKTIVLYPPLMKEYQEIFSDSGESDDEAFIYKPDIHEK-KGKPKEAEDEKDEKS--KKNKTPIDDIKDWWSNIKGEWHL

RBUV\_G --NTSRIYYPPLSKEYESVFSDSGEMEDEAYIYKPEIKHKGNVNPKKDDEDSSEEE--KKNKTPLDDLSDWWKKLKGEWKL

IRIRV\_G --NVSYIYYPPLIKEYEEMFSDSGEIEDEAFIFQPEIHSK-KTKPKSDDQDSSTDS--KKNKTPIDDLEDWWKKLKSEWGA

ITAV\_G -PGTWVISVPDSRQDLEDILSWEHLEGERVVYDERVILTSSDIELDDLGNTLDTEV---ETKNIFSGIIKWFDSVYSS---

\* : .. .. \*

**Transmembrane domain**

VSIV\_G -IASFFFIIGLIIGLFLVLRVSIYLCIKLKHTKK--------RQIYTDIEMNRLGK-----------------------

**XIBV\_G** -WSSGIQAVISIFFIFVFVAALYKLTQICIRIKR--------RIHPSKKKIQSPDPEADADLIGSKPERDNRNQTVFG-

**ARUV\_G** -WTSGIKAVLGTAVLIVMVVILWRVVMICKAIKC--------KSGNASKTNDSAES-----LFVPKTS-DHRLAKMFG-

**INHV\_G** -WTSGGKSFILMILMLCSGYLIIKLSLFCYLKVK--------RRKSIPDGMETKKRSKQNEYVDLKPLKNGSRNNPFSV

**SBAV\_G** GFKSLLAWVTTVIHISIIVVCILIIVKLYKTCKRNQSQATHATKRTGKMLDTIIGEDHLYDTVKTPSRNEDIRLNLFNH

CURV\_G -IKGILIGLFTFALLIGVVKLGVFIKSSFRKRRDDSIPEGKDEEIGIKMQSRRSRQNIYEEINEVSPTMTRRGRNIFN-

RBUV\_G -IKGIIVSFLVFLGLYLIIKCCLKLRSVIKEKK---IKKVVDEEKSLELKERRAEPNIYEEINETSRPRVRRGRNYFN-

IRIRV\_G -IKGTLIGVTLIVGLILLLKLILAVKRLISRKRG-KKSEAGKEEKEKEMVRFRATPNIYEEIDERSPTRPRGRRTIFT-

ITAV\_G -VSSYIYIIGVLVAVGLGIWILTKAKRFIVRDRS-----PGNQTSEITIPMNRNRREDQRSYI----------------

. :

**Figure 3.** A Clustal X alignment of the G proteins of VSIV, curioviruses and the proposed arurhaviruses (indicated in bold). Conserved cysteine residues in the VSIV G protein are marked (CI-CXII) as are other pairs of cysteine residues that may form additional disulphide bonds (a-f). Cysteine residues in the extended N-terminal domains of curioviruses are also shown.



**Figure 4.** The evolutionary history was inferred from a Clustal W alignment of 128 complete L protein sequences of 124 animal rhabdoviruses currently assigned or recently proposed for assignment to species, and four proposed members of the genus *Arurhavirus* (ARUV, XIBV, RBUV, ITAV). Phylogenetically informative sites were selected from the alignment using Gblocks resulting in 1019 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 (-112482.5480) 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 arurhavirus N proteins.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ARUV | XIBV | INHV | SBAV |
| ARUV |  |  |  |  |
| XIBV | 67.4 |  |  |  |
| INHV | 57.2 | 58.4 |  |  |
| SBAV | 43.7 | 42.1 | 44.0 |  |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ARUV | XIBV | INHV | SBAV |
| ARUV |  |  |  |  |
| XIBV | 39.3 |  |  |  |
| INHV | 36.0 | 39.1 |  |  |
| SBAV | 27.0 | 23.8 | 27.1 |  |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ARUV | XIBV | INHV | SBAV |
| ARUV |  |  |  |  |
| XIBV | 66.5 |  |  |  |
| INHV | 63.3 | 61.7 |  |  |
| SBAV | 50.7 | 50.4 | 50.0 |  |