

Genomic and phylogenetic characterization of Shuni virus

Charmaine van Eeden, Frank Harders, Jeroen Kortekaas, Alex Bossers, Marietjie Venter*

Abstract Shuni virus (SHUV), a member of the genus *Orthobunyavirus*, has in a recent study been associated with neurological disease in horses in South Africa. After its first isolation in 1966 from an asymptomatic bovine, very little attention was given to the genetic characterisation of SHUV. The association of SHUV with severe neurological disease in several horses in South Africa prompted us to determine the full genome sequence of a horse neurovirulent isolate to compare it to other members of the genus *Orthobunyavirus*, as well as the partially sequenced genome of the prototype SHUV strain. The availability of a full genome sequence will facilitate the development of a reverse genetics system to study SHUV molecular biology and pathogenesis.

Accession numbers: The GenBank accession numbers for SAE 18/09 are KC510272 for the S segment, KF153117 for the M segment and KF153118 for the L segment.

C. van Eeden · M. Venter (*)
Zoonoses Research Unit, Department of Medical Virology,
University of Pretoria, Pretoria, South Africa
e-mail: marietjie.venter@up.ac.za

C. van Eeden
e-mail: charmaine.vaneeden@gmail.com

F. Harders · J. Kortekaas · A. Bossers
Central Veterinary Institute, Part of Wageningen University and
Research Centre, Wageningen, The Netherlands
e-mail: frank.harders@wur.nl

Introduction

Shuni virus (SHUV) belongs to the Simbu serogroup of the genus *Orthobunyavirus*, family *Bunyaviridae*. Given the importance of the orthobunyaviruses assigned to the Simbu serogroup in both public and veterinary health, surprisingly little genetic information is available for SHUV, with limited clinical descriptions of disease dating back to the 1960s and 1970s, including isolation of the virus from a febrile child [1]. The prototype SHUV strain (An10107) was isolated during arbovirus surveys at an abattoir in Nigeria in 1966 from an apparently healthy cow [2]. In 2012, an almost complete genome sequence of this strain was submitted to the GenBank database as part of a phylogenetic analysis of the Simbu serogroup [3].

The identification of highly pathogenic strains of SHUV from the brains of horses with severe neurological disease in South Africa [1] raises several questions regarding the difference in pathogenicity that full genome sequencing may address. Although not yet fully characterised, the prototype sequence gave us the opportunity to compare SAE18/09, an isolate that had been associated with neurological disease in an equine [4], to this isolate from an asymptomatic bovine [2]. Factors such as spatial and

J. Kortekaas
e-mail: jeroen.kortekaas@wur.nl

A. Bossers
e-mail: alex.bossers@wur.nl

Present Address:

M. Venter
Division of Global Health Protection, US Centres for Disease
Control and Prevention, Pretoria, South Africa

temporal separation as well as possible recombination events or point mutations could affect the genotypic and phenotypic traits resulting in the observed pathogenic differences between these two strains, and thus the need to investigate the molecular biology of SHUV arose.

Orthobunyaviruses are enveloped and possess a genome composed of three segments of negative-sense single-stranded RNA, designated as large (L), medium (M) and small (S) [5]. The L segment encodes the viral RNA polymerase, which functions in replication and transcription of the genomic RNA segments. The M segment codes for a glycoprotein precursor, which is cotranslationally cleaved into the viral envelope glycoproteins, Gn and Gc, as well as a non-structural protein, named NSm [5, 6]. The S segment encodes two proteins, the nucleocapsid (N) protein and a smaller non-structural protein (NSs), in overlapping reading frames [5, 7]. A characteristic feature of orthobunyavirus genome segments is the complementarity of the 3' and 5' termini, the base pairing of which results in so-called panhandle structures [5].

Alignments of S segments of viruses from the Simbu serogroup have revealed nucleotide sequence identities between 65 % and 99.1 % [8]. The nucleotide sequence identity of the M segment open reading frame (ORF) ranges from 47.9 % to 56.0 % between members of the Simbu serogroup. The amino acid (aa) sequences are highly variable (37.3 %-49.7 %), although several features are conserved. These features include 59 of the 71 cysteine residues and a conserved arginine (R) residue, which is located proximal to the predicted cleavage site between Gn and NSm [9].

Analysis of the L protein has shown the nucleotide sequence identity between members of the Simbu serogroup to range between 60.4 % and 92.9 %. The conservation at the aa level varies from 57.5 % to 98.4 % [3]. Four conserved regions have been identified in the L protein [10].

As in the case of other viruses that have segmented genomes, the occurrence of genetic reassortments of orthobunyaviruses has been reported. Ngari virus (MRIV), for instance, was generated as a result of reassortment between BUNV and Batai viruses (BATV), resulting in increased pathogenicity and outbreaks of haemorrhagic fever [11, 12]. For the Simbu serogroup, the Jatobal (JATV) and Tinaroo (TINV) viruses are likely reassortants containing RNA segments from OROV and AKAV, respectively [13, 14]. Reassortment has also been reported to occur between Australian field isolates of Aino (AINOV) and Peaton (PEAV) viruses [15], which are closely related to SHUV both serologically and phylogenetically.

The aim of this study was to provide a detailed genomic and phylogenetic characterisation of SHUV (SAE18/09) and to compare the sequence of this pathogenic strain to that determined for the prototypic strain (An10107).

Materials and methods

Virus isolate

SAE 18/09, originally isolated from the brain of a horse, was cultured on Vero cells in EMEM containing 2 % foetal calf serum (Gibco BRL, Invitrogen). Cultures were incubated at 37 °C until 70-80 % cytopathic effect was exhibited and harvested.

Genome sequencing

Forty-five milliliters of SAE 18/09 cell culture supernatant was put through Amicon Ultra 10 K centrifugal filters (Merck, Darmstadt, Germany). RNA was isolated from the concentrate using TRIzol LS (Sigma-Aldrich, Missouri, United States). The extracted material was purified using an RNA Clean and Concentrator-5 kit (Zymo Research, California, United States); the in-column DNase step was included. The sample was prepared for shipping using the GenTegra RNA system (IntegenX, California, USA). Libraries were prepared using a ScriptSeq2 RNA-Seq Library Preparation Kit (Epicentre Biotechnologies, Wisconsin, United States). Both the protocol for FFPE samples and the protocol for alternative fragmentation (2 min at 85 °C instead of 5 min) were utilized according to manufacturer's instructions unless otherwise stated. The quality of the libraries was evaluated using a Bioanalyzer with a High Sensitivity DNA Kit (Agilent Technologies, California, United States), and the quantity was determined using a Qubit dsDNA HS Assay Kit (Life Technologies, New York, United States). Cluster generation and paired-end 250-bp sequencing of the libraries was done on an Illumina MiSeq V2 instrument. Raw sequencing data were first trimmed and QC filtered using in-house scripts (<http://galaxy.wur.nl>). Possible contaminants were identified (host cell fragments) by short-read mapping using Bowtie2 [16] and subsequent filtering.

Full *de novo* genome reconstruction was achieved by sequence assembly using ABySS (version 1.3.3; Canada's Michael Smith Genome Science Centre). Repeated short-read mapping using Bowtie2 and manual inspection of any unusual mapping densities were done to reduce errors in the assembly process. Finally, the full genome was reconstructed using closely related reference sequences, where we determined the synteny of the initial contigs from ABySS. A mapping-based approach in which sequence reads were mapped using Bowtie2 on the sequence fragments of the three segments determined earlier was used to confirm or refine earlier derived sequences. In addition, this allowed the reconstruction of the full sequences of the 5' and 3' ends of all three segments. Further manual curation by both complementary approaches into a single high-

quality sequence was performed using SeqMan (DNA-STAR Lasergene version 11). Sequences were verified by classical Sanger sequencing.

Phylogenetic and sequence analysis

Nucleotide phylogenetic analysis was carried out using the Tamura-Nei model and amino acid analysis using the Jones-Taylor-Thornton (JTT) model [17], both of which are included in MEGA version 5, using 1000 bootstrap analysis [18]. Alignments were carried out using the ClustalW subroutine [19], which is included in the Bioedit program. Sequence similarity was determined using the distance estimation program of MEGA v5 [18], using the p-distance model [20]. Analysis of cleavage sites, glycosylation sites and transmembrane domains was done using the following programs: SignalP (<http://www.cbs.dtu.dk/services/SignalP>), NetNgly (<http://www.cbs.dtu.dk/services/NetNGlyc>) and TMHMM (<http://www.cbs.dtu.dk/services/TMHMM>).

Results

Genome segment sequence analysis

The full genome sequence of SAE 18/09, isolated from the brain of a horse that died with severe neurological signs [1], was determined, and its conserved features and domains were compared to those of available Simbu serogroup viruses.

S segment

The S segment of SAE 18/09 was found to be 850 nucleotides (nt) in length, with a 3' untranslated region (UTR) of 33 and a 5' UTR of 115 nucleotides (Fig. 1a, Table 1). Phylogenetically, the S segment of SAE 18/09 groups with Aino (AINOV) and Kaikular (KAIV) viruses (Fig. 2a), with percentage identities of 93.9 % (AINOV) and 93.7 % (KAIV), respectively (Table 1). Certain regions of the N protein are conserved amongst members of the genus *Orthobunyavirus*, and these six domains were well conserved between SAE 18/09 and other members of the Simbu serogroup, as was observed by Saeed et al. [14]. Of significance is the observation that within the first conserved domain, the first Shuni isolate from Nigeria (AF362405) had an asparagine (N) residue at position 9 while all other members of the serogroup had an aspartic acid (D) residue at this position. The four residues involved in the formation of the ribonucleoprotein complexes (P₁₂₅, G₁₃₁, Y₁₅₈ and I₂₃₁) [21] were all conserved in SAE 18/09 (Fig. 1a). In the NSs, which has been shown to play a key

role in pathogenesis ([22–24], a single non-synonymous codon difference (L₈₄) was observed between the two SHUV isolates.

M segment

The M segment of SAE 18/09 was 4351 nt in length, with a 3' UTR of 23 and a 5' UTR of 113 nucleotides (Fig. 1b). Phylogenetic analysis of the M segment sequences of members of the Simbu group demonstrated that SAE 18/09 was most closely related to Schmallenberg virus (SBV), with 61.4 % nucleotide sequence identity (Fig. 2b). It had 51.8 % nucleotide sequence identity to JATV and 49.4 % to SHAV (Table 1). The predicted signal peptide of the SAE 18/09 glycoprotein precursor as determined by SignalP 4.1, was situated between aa 17 and 18 (GVP-IP). The Gn/NSm cleavage site KSLRxAR, which is conserved amongst members of the California and Bunyamwera serogroups [6, 25–27] (Fig. S1), is less well conserved within the Simbu serogroup. Based on alignment with these viruses and comparison to the data obtained by Yanase et al. [9], protease cleavage between Gn/NSm of SAE18/09 occurs after the conserved arginine (R) 308 residue (Fig. S1).

The cleavage site between NSm and Gc is poorly conserved amongst members of the Simbu serogroup [9], though between other members of the genus *Orthobunyavirus*, cleavage usually follows a conserved alanine residue [26–28]. SignalP predicts cleavage of SAE18/09 at VDA₄₆₄-ND (Fig. S1). Based on these analyses, the Gn protein comprises 291aa, NSm 156 aa, and Gc 940 aa – sizes that correlate well with what was found previously for other Simbu viruses [28]. However, SHUV isolate An 10107 differed at three positions with SAE18/09 at the NSm/Gc cleavage site, resulting in a predicted cleavage at AAS₄₆₅-DK, changing the size of the NSm to 157 aa and that of Gc to 939 aa.

The number and positions of N-linked glycosylation sites (N-X-S/T) are not well conserved amongst members of the Simbu serogroup [9], and in this analysis using NetNGlyc 1.0, four potential glycosylation sites were identified in the glycoprotein precursor of both SAE 18/09 and An10107: two in Gn and two in Gc (Fig. 1b, Fig. S1). The site at the amino terminus of Gn (N₄₀) is well conserved amongst members of the Simbu serogroup, as is N₄₉₂, at the amino terminus of Gc. Our analysis showed 60 of 71 cysteine residues to be conserved between all Simbu serogroup viruses analysed, which was one more than originally found by Yanase et al. [9] (Fig. 1b).

Transmembrane domains were determined using TMHMM. Two regions are identified in Gn (212-229) and (234-253), which together form a very long hydrophobic sequence (212-253) that is followed by charged amino acids (254-267), similar to stop-transfer sequences seen in the transmembrane domains of other viral envelope

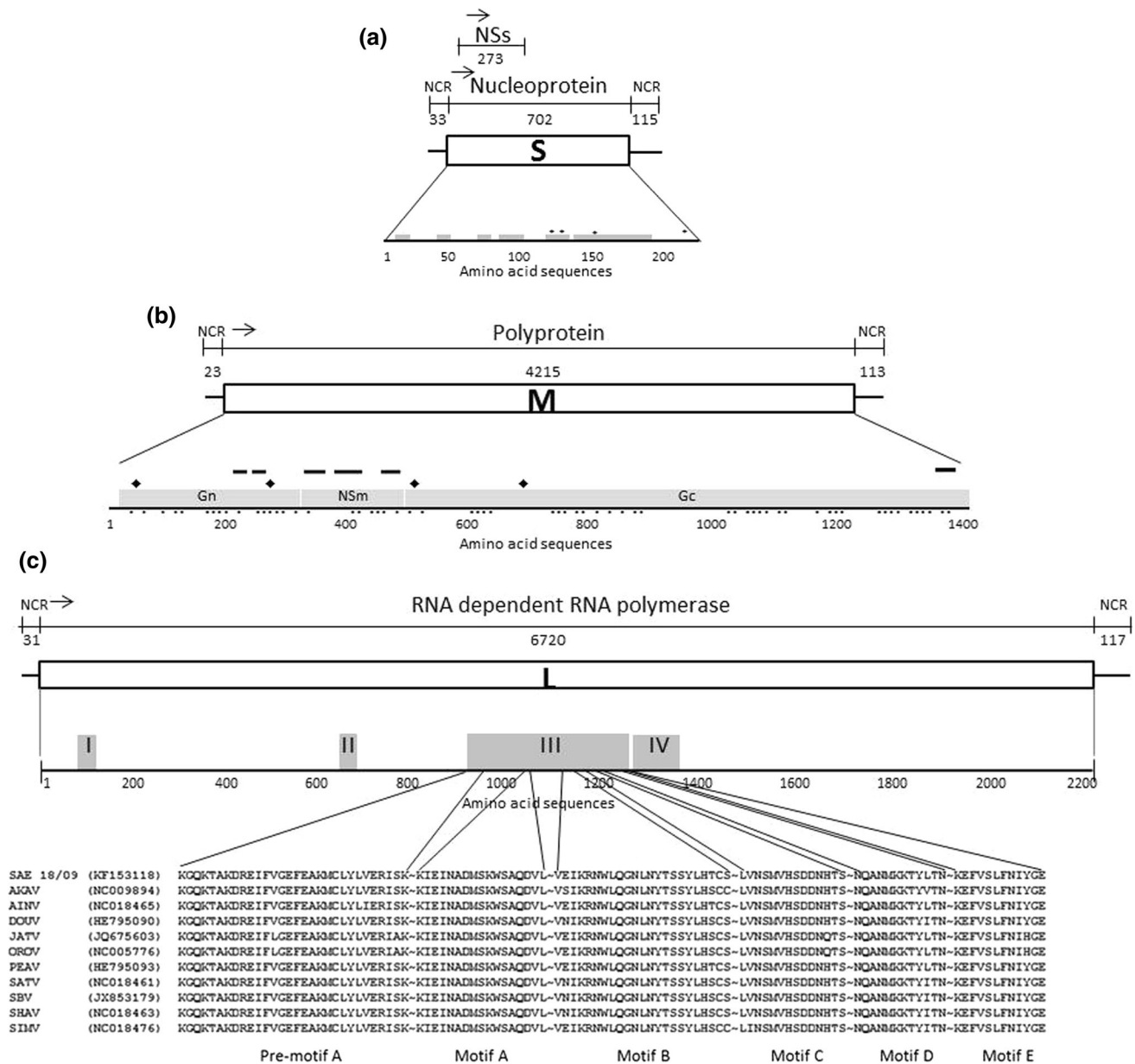


Fig. 1 Genome characteristics of Shuni virus (SAE 18/09). (a) The S segment encodes the N and NSs proteins in overlapping reading frames. The grey boxes indicate the conserved regions identified by Saeed et al. [14], and the black diamonds indicate the four residues involved in the formation of the ribonucleoprotein complexes. (b) The M segment encodes the GPC, which is cotranslationally cleaved into

proteins [5, 6]. In NSm, three regions were identified (315-337, 371-393, 445-462); and in Gc, one (1357-1379) (Fig. 1b, Fig. S1). The latter likely functions as a membrane anchor [6, 29].

L segment

The L segment of SAE 18/09 is 6910 nt in length with a 31-nt 3' UTR and a 117-nt 5' UTR (Fig. 1c). This is

comparable with other members of the Simbu serogroup, although SAE 18/09 has a significantly longer 5' UTR. Only three complete Simbu group L segments were available for phylogenetic analysis. SAE 18/09 branched separately from the other Simbu serogroup members (Fig. 2c). In terms of nucleotide sequence identity, the complete L segment of SAE 18/09 was most similar to SBV (66.2 %), followed by AKAV (65.8 %) and OROV (61.8 %) (Table 1). A comparison of the L protein aa

Table 1 Characteristics of the S, M and L segments of five Simbu serogroup viruses compared with SAE 18/09

Virus	S Segment								
	Nucleotide sequence					Amino acid sequence			
	Total length*	3' UTR	5' UTR	N*	NSs*	N‡	NSs‡		
SAE18/09 (KC510272)	850	33	115	702	276	233	91		
SHUV (HE800143)	–	–	–	702 (97.9)	276 (99.3)	233 (100)	91 (98.9)		
KAIV (AF362394)	850 (93.7)	34	114	702 (93.6)	276 (99.3)	233 (98.3)	91 (98.9)		
AINOV (HE795089)	850 (93.9)	34	114	702 (94)	276 (98.6)	233 (99.6)	91 (96.7)		
PEAV (HE795095)	–	–	–	702 (91.7)	276 (97.5)	233 (99.6)	91 (94.5)		
SIMV (HE795110)	867 (77.5)	33	132	702 (72.2)	276 (83)	233 (79.8)	91 (62.7)		
Virus	M Segment								
	Nucleotide sequence						Amino acid sequence		
	Total length*	3' UTR	5' UTR	Gn*	NSm*	Gc*	Gn‡	NSm‡	Gc‡
SAE18/09 (KF153117)	4351	23	113	873	468	2820	291	156	940
SHUV (HE800142)	–	–	–	873 (99.2)	471 (97.9)	2817 (98.8)	291 (98.7)	157 (96.2)	939 (97.9)
AINOV (HE795088)	–	–	–	873 (92.5)	468 (70.3)	2820 (83.7)	291 (89.3)	156 (59)	940 (74.4)
SATV (HE795103)	–	–	–	873 (82.9)	468 (56.7)	2817 (68.4)	291 (73.4)	156 (39.3)	939 (53.8)
SBV (HE649913)	–	–	–	873 (82.7)	468 (57.6)	2817 (68.3)	291 (72.7)	156 (42.7)	939 (53.1)
SIMV (HE795109)	–	–	–	873 (78.5)	486 (57.4)	2814 (64.9)	291 (66.6)	162 (40.4)	938 (48.2)
Virus	L Segment								
	Nucleotide sequence				Amino acid sequence				
	Total length*	3' UTR	5' UTR	RdRp*	RdRp‡				
SAE18/09 (KF153118)	6910	31	117	6762	2253				
SHUV (HE800141)	–	–	–	6762 (93.5)	2253 (98.9)				
AINOV (HE795087)	–	–	–	6762 (85.1)	2253 (95.2)				
PEAV (HE795093)	–	–	–	6759 (78.1)	2252 (88.9)				
SANV (HE795099)	–	–	–	6762 (78.7)	2253 (90.5)				
SIMV (NC018476)	–	–	–	6762 (67.9)	2253 (70.9)				

* Values are lengths in nucleotides (percentage identity to SAE 18/09)

‡ Values are lengths in amino acids (percentage identity to SAE 18/09)

_ Not available

sequences of SAE 18/09 with those of other Simbu serogroup members showed homology at the endonuclease active site H...D...PD....DxK.....T at positions 34, 53, 76, 77, 90, 92 and 103 respectively [30]. Conserved regions I and II, containing the conserved residues PD (aa 76-77) and RY (aa 650-651), respectively, were also identified in all members investigated. Pre-motif A as well as motifs A, B, C, D and E, which make up region III (aa 947-1238), are highly conserved, with the only significant difference being that AINOV has a V → I₉₇₁ substitution, which is not shared by any other members, including SAE 18/09 (Fig. 1c). The four conserved residues of region IV – G₁₂₇₄, Y₁₂₉₄, G₁₃₂₃ and G₁₃₃₆ [10] – showed 100 % homology.

Phylogeny of the viral proteins

Since few full-genome sequences of members of the Simbu serogroup were available, phylogenetic analysis was also carried out for the individual viral proteins to determine whether any reassortment had occurred. All protein sequence comparisons resulted in clustering of SAE 18/09 with the prototype SHUV isolate (An10107), suggesting that no reassortment had occurred relative to the two strains that would account for the varied pathogenicity observed. Shuni virus showed the same relative clustering to the members of the orthobunyavirus serogroup for all proteins, which does not suggest reassortment.

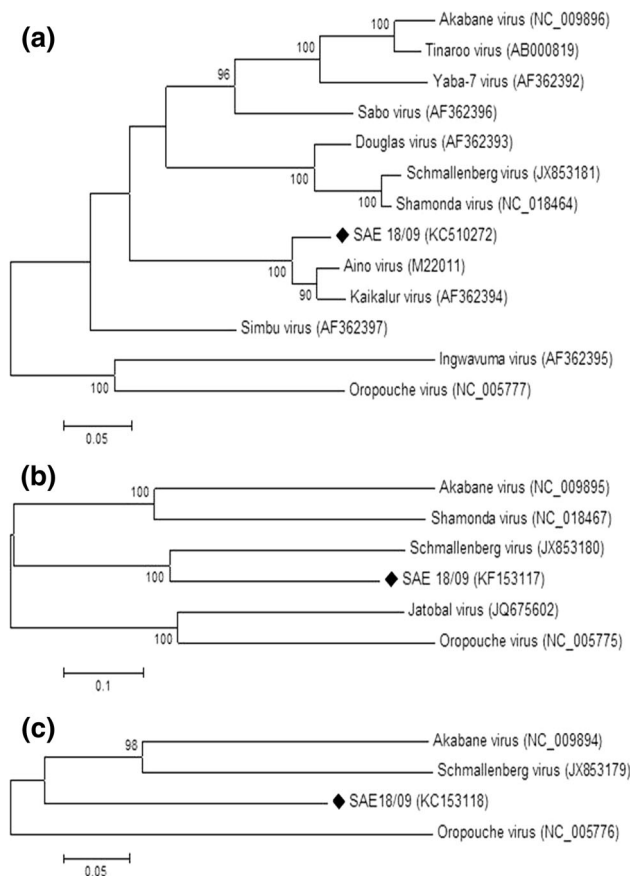


Fig. 2 Phylogenetic relationship between the S (a), M (b) and L (c) segments of SHUV (SAE 18/09) and those of other available Simbu serogroup virus sequences. Maximum-likelihood trees were constructed using the Tamura-Nei substitution model in the MEGA 5 program with representative full genome sequences of other Simbu serogroup viruses. Estimates were based on bootstrap resampling carried out with 1000 replicates and drawn to scale with bars indicating 0.05, 0.1 and .05 nucleotide substitutions, respectively. The GenBank accession numbers of reference strains are indicated

Nucleoprotein (N)

The N gene was found to comprise 702 nt coding for 233 amino acids. The N protein was analysed at both the nt and aa levels. At the nucleotide level, SAE 18/09 grouped with the prototype Shuni virus strain (An 10107) (97.9%). In the Aino/Kaikalar clade, the grouping was similar to that obtained with the whole S segment. At the amino acid level, however, SAE18/09 grouped with Shuni, Peaton and Sango viruses, but the bootstrap values for this branching was low (55%) (Fig. 3a). Distance analysis revealed the sequence identity between the two SHUV isolates to be 100%, with interspecies values of 99.6% (Peaton, PEAV), 99.6% (AINOV), 99.6% (Sango, SANV) and 98.3% (KAIV) (Table 1).

Non-structural protein S segment (NSs)

The NSs gene consists of 276 nt; and the encoded protein, 91 aa. Phylogenetic analysis of NSs revealed similar groupings at both the nt and the aa level. SAE 18/09 grouped with SHUV (An 10107), AINOV, KAIV, and SANV, although bootstrap support was low. Analysis of NSs at the nt level showed 99.3% identity between the SHUV isolates, and interspecies values of 99.3% (KAIV) and 98.6% (AINOV). Amino acid analysis gave a similar result, with a value of 98.9% between the SHUV isolates and interspecies values of 98.9% (KAIV) and 96.7% (AINOV) (Fig. 3e).

Glycoprotein precursor (GPC)

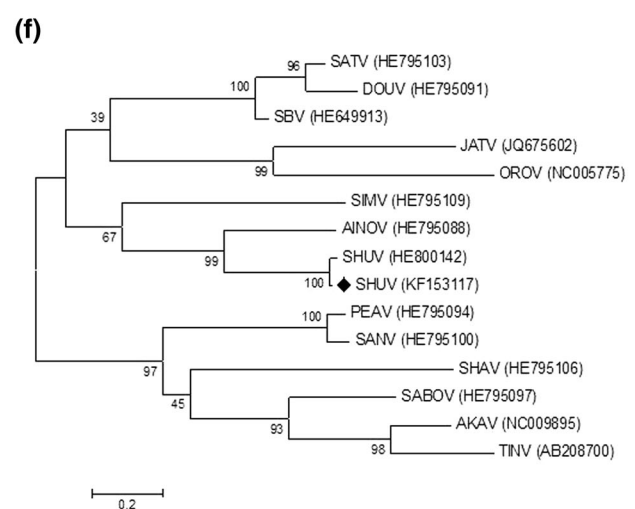
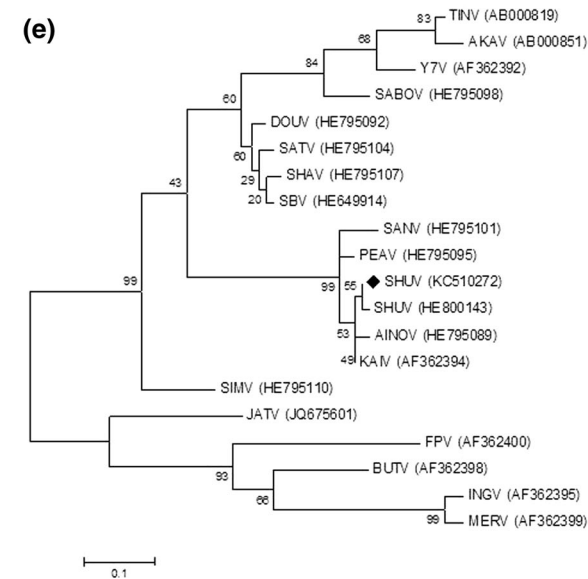
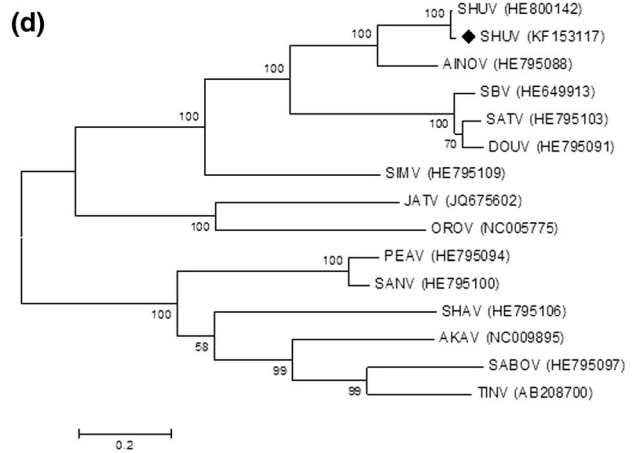
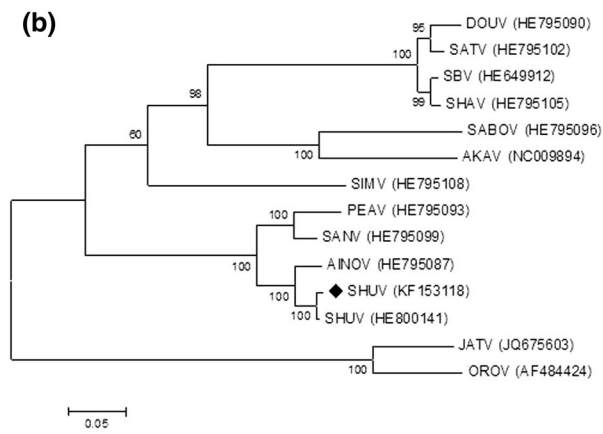
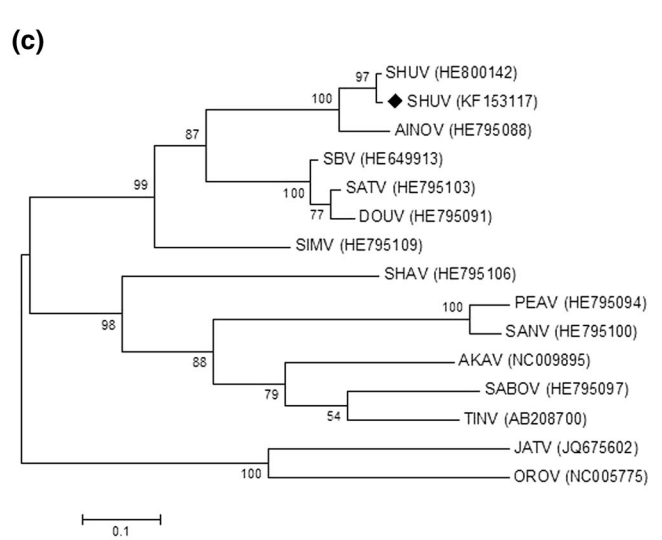
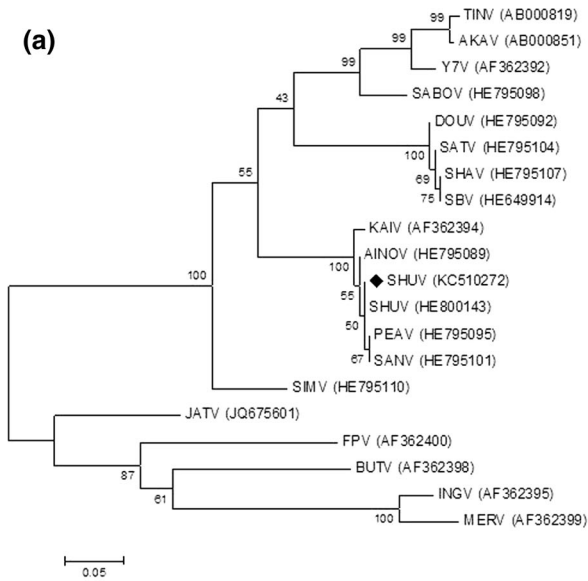
The M segment codes for a 1404-aa protein. Phylogenetic analysis revealed that SAE 18/09 groups with SHUV and AINOV at both the aa and nt level, with 100% bootstrap support for the complete GPC. Distance analysis at the nt level showed identity values of 98.8% between the SHUV isolates and interspecies values of 84.9% (AINOV) and 70.7% (SATV). At the aa level, the SHUV isolates shared 97.9% identity. The interspecies values were much lower, with 75.9% (AINOV) and 56.4% (SATV), reflecting the low level of conservation in the orthobunyavirus GPC.

Gn

Amino acid (Fig. 3c) and nucleotide sequence analysis of the Gn protein of SAE 18/09 resulted in the same topology as the full GPC analyses, where SAE 18/09 grouped with SHUV and AINOV within a clade including DOUV, SATV and SBV. Distance estimation showed the Gn protein to be more conserved than the precursor, with 99.2% nucleotide sequence identity between the SHUV isolates, 92.5% to AINOV and 82.9% to SATV. The amino acid identity value between the SHUV isolates was 98.7%, followed by interspecies values of 89.3% (AINOV) and 73.4% (SATV).

NSm

Analysis of the NSm sequence revealed a topology similar to that of Gn. The genetic distance between the Simbu group viruses is low for NSm [6, 9], and in the nt analysis SAE 18/09 was 97.9% identical to SHUV (An 10107) and only 70.3% similar to AINOV. At the aa level, the two SHUV isolates shared 96.2% identity, followed by 59% identity to AINOV (Fig. 3f).



◀**Fig. 3** Relationship of Shuni virus (SAE 18/09) to other members of the Simbu serogroup. Maximum-likelihood trees of the amino acid sequences of the N (a), L (b), Gn (c), Gc (d), NSs (e) and NSm (f) proteins were constructed using the Jones-Taylor-Thornton (JTT) substitution model in the MEGA 5 program. Estimates were based on bootstrap resampling carried out with 1000 replicates and drawn to scale with bars indicating 0.05, 0.05, 0.1, 0.2, 0.1 and 0.2 nucleotide substitutions, respectively. The GenBank accession numbers of reference strains are indicated

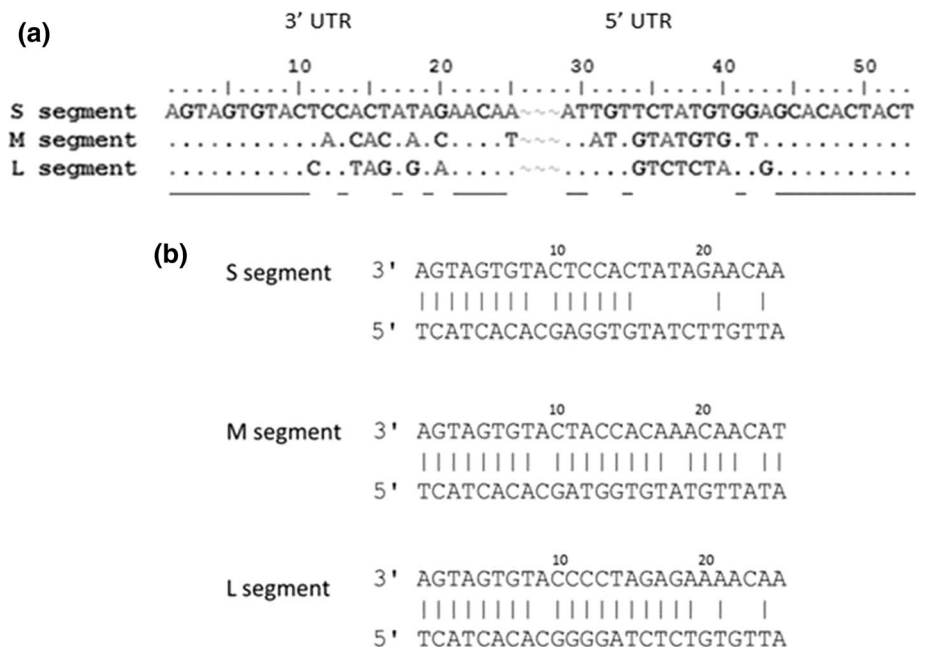
Gc

Phylogenetic analysis of Gc resulted in the same topology as was observed for Gn at both the nt and the aa level. Nucleotide sequence identity between the two SHUV isolates was 98.8 %, with an interspecies identity of 83.7 to AINOV. Amino acid analysis showed 97.9 % identity between the SHUV isolates and 74.4 % identity to AINOV. The lower level of homology in the Gc protein is well documented for the orthobunyaviruses [5, 6, 9] (Fig. 3d).

RNA-dependent RNA polymerase (RdRp)

The RdRp comprises 2253 aa. Phylogenetic analysis of the RdRp protein of SAE 18/09 grouped it together with SHUV (An 10107) in a clade with AINOV, with high bootstrap support at both the nt and the aa level (Fig. 3b). Sequence identity at the nt level demonstrated SAE 18/09 to be 93.5 % similar to SHUV, 85.1 % to AINOV, and 78.7 % to SANV. At the aa level, it was found to be 98.9 % similar to SHUV, 95.2 % to AINOV and 90.5 % to SANV.

Fig. 4 Sequence comparisons to identify highly conserved nucleotides in the non-coding regions of the S, M and L segments of SAE 18/09, with conserved nucleotides underlined (a). Analysis of the complementarity between the 5' and 3' non-coding regions of the S, M and L segments of SAE 18/09 (b)



Terminal sequences

The 5' untranslated region (UTR) sequences of all three SAE 18/09 genomic segments are identical for the first 10 nt, as are the corresponding 3' UTR sequences (Fig. 4a). In addition, these 5' and 3' sequences exhibit complementarity for all but one of the first 11 nt; the mismatch at position 9 is conserved amongst members of the *Bunyaviridae* [5, 31, 32]. Complementarity extended to 15 nt in the S segment, 17 nt in the M segment, and 19 nt in the L segment (Fig. 4b), which was similar to what has been seen in Bunyamwera virus, where complementarity extends to 15, 18 and 19 nt, respectively [33, 34]. The AACAA motif, which is associated with minireplicon activity [35], was found to be conserved for all three segments of SAE 18/09 (AACAA₂₄).

Discussion

The genome sequence of the pathogenic SHUV strain SAE18/09 was determined and compared with the prototype strain An10107, isolated from an asymptomatic bovine. Though highly similar, distinct differences were found between the sequences of these two strains.

Only a limited number of viral sequences were available for the full segment analyses; however, the overall topology for the S, M and L segments was uniform and reflected the results seen in other partial genome studies [3, 8]. The analysis of the viral proteins was more comprehensive, and for all the proteins, SHUV SAE18/09 grouped with SHUV

An10107 and AINOV, its closest serological relative, the topology reflecting the original molecular analysis of the Simbu serogroup viruses [8]. Sequence identity between the two SHUV isolates was high and ranged between 93.5 and 99.3 % at the nt level and 96.2-100 % at the aa level for all viral proteins. In particular, for the N ORF, the values of 97.9 % (nt) and 100 % (aa) were very high for isolates from different geographic locations and isolated decades apart but reflected what has been seen in other bunyaviruses such as RVF. However, for Aino and Akabane viruses, the identity between Australian and Japanese strains can result in differences up to 7-8.9 % [36, 37]. The highest variation existed in the NSm and Gc proteins, Gc is exposed on the outer surface of the virion and is thus a principal target of neutralizing antibodies. Such variation may play a role in positive selection, resulting in drift and potential immune evasion over time.

A noteworthy variation that has been linked with differences in pathogenicity in other viruses was that, in the prototype SHUV isolate (An10107), the predicted cleavage site for NSm/Gc produced an NSm and a Gc protein one aa longer and shorter, respectively, due to three aa changes at the cleavage site. The significance of these changes will need to be investigated in studies of expression and pathogenicity, as they can lead to changes in the efficiency of signal peptide cleavage, which in turn can affect the production of infectious virions [38, 39]. The high level of conservation between these two isolates – one of neurologic origin and the other asymptomatic origin – further highlights the need to investigate the factors that influence SHUV pathogenicity.

In summary, the genome of SHUV shared all of the features and conserved domains previously identified for the Simbu serogroup. Phylogeny of the viral proteins was similar throughout, with no indication of genomic reassortment. This genome sequence and phylogenetic analysis marks an important step towards the establishment of a reverse genetics system to analyse the biology and pathogenicity of SHUV, as well as towards the further elucidation of the molecular relationship between members of the Simbu serogroup. Nevertheless, more full-genome sequences are required to fully elucidate the relationship between the Simbu serogroup viruses, and the mutations observed between the two SHUV strains at the NSm/Gc cleavage site need to be investigated further.

Acknowledgments This study was funded by the Global Disease Detection (GDD) programme of the Centres for Disease Control and Prevention (CDC), Atlanta, USA.

References

- Moore DL et al (1975) Arthropod-borne viral infections in man in Nigeria: 1964-1970. *Ann Trop Med Parasitol* 69:49-64
- Causey OR et al (1972) Isolations of Simbu-group viruses in Ibadan, Nigeria 1964-69, including the new types Sango, Shamonda, Sabo and Shuni. *Ann Trop Med Parasitol* 66(3):357-362
- Goller KV et al (2012) Schmallerberg virus as possible ancestor of Shamonda virus. *Emerg Infect Dis* 18(10):1644-1646
- van Eeden C et al (2012) Shuni virus as cause of neurologic disease in horses. *Emerg Infect Dis* 18(2):318-321
- Elliott RM (1990) Molecular biology of the Bunyaviridae. *J Gen Virol* 71(Pt 3):501-522
- Fazakerley JK et al (1988) Organization of the middle RNA segment of snowshoe hare Bunyavirus. *Virology* 167(2):422-432
- Fuller F, Bhowan AS, Bishop DH (1983) Bunyavirus nucleoprotein, N, and a non-structural protein, NSS, are coded by overlapping reading frames in the S RNA. *J Gen Virol* 64(Pt 8):1705-1714
- Saeed MF et al (2001) Phylogeny of the Simbu serogroup of the genus Bunyavirus. *J Gen Virol* 82:2173-2181
- Yanase T et al (2003) Sequence analysis of the medium RNA segment of three Simbu serogroup viruses, Akabane, Aino, and Peaton viruses. *Virus Res* 93(1):63-69
- Ogawa Y et al (2007) Sequence determination and functional analysis of the Akabane virus (family Bunyaviridae) L RNA segment. *Arch Virol* 152(5):971-979
- Briese T et al (2006) Batai and Ngari viruses: M segment reassortment and association with severe febrile disease outbreaks in East Africa. *J Virol* 80(11):5627-5630
- Yanase T et al (2006) Genetic characterization of Batai virus indicates a genomic reassortment between orthobunyaviruses in nature. *Arch Virol* 151(11):2253-2260
- Kobayashi T et al (2007) Genetic diversity and reassortments among Akabane virus field isolates. *Virus Res* 130(1-2):162-171
- Saeed MF et al (2001) Jatobal virus is a reassortant containing the small RNA of *Oropouche virus*. *Virus Res* 77(1):25-30
- Yanase T et al (2010) Genetic characterization of Aino and Peaton virus field isolates reveals a genetic reassortment between these viruses in nature. *Virus Res* 153(1):1-7
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275-282
- Tamura K et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-4680
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5):418-426
- Eifan SA, Elliott RM (2009) Mutational analysis of the Bunyamwera orthobunyavirus nucleocapsid protein gene. *J Virol* 83(21):11307-11317
- Bridgen A et al (2001) Bunyamwera bunyavirus nonstructural protein NSs is a nonessential gene product that contributes to viral pathogenesis. *Proc Natl Acad Sci USA* 98(2):664-669
- Muller R et al (1995) Characterization of clone 13, a naturally attenuated avirulent isolate of Rift Valley fever virus, which is altered in the small segment. *Am J Trop Med Hyg* 53(4):405-411
- Varela M et al (2013) Schmallerberg virus pathogenesis, tropism and interaction with the innate immune system of the host. *PLoS Pathog* 9(1):e1003133
- Briese T, Kapoor V, Lipkin WI (2007) Natural M-segment reassortment in Potosi and Main Drain viruses: implications for the evolution of orthobunyaviruses. *Arch Virol* 152(12):2237-2247

26. Briese T, Rambaut A, Lipkin WI (2004) Analysis of the medium (M) segment sequence of Guaroa virus and its comparison to other orthobunyaviruses. *J Gen Virol* 85(Pt 10):3071–3077
27. Pollitt E et al (2006) Characterization of Maguari orthobunyavirus mutants suggests the nonstructural protein NSm is not essential for growth in tissue culture. *Virology* 348(1):224–232
28. Savji N et al (2011) Genomic and phylogenetic characterization of Leanyer virus, a novel orthobunyavirus isolated in northern Australia. *J Gen Virol* 92(Pt 7):1676–1687
29. Pekosz A et al (1995) Protection from La Crosse virus encephalitis with recombinant glycoproteins: role of neutralizing anti-G1 antibodies. *J Virol* 69(6):3475–3481
30. Reguera J, Weber F, Cusack S (2010) Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathog* 6(9):e1001101
31. Elliott RM, Schmaljohn CS, Collett MS (1991) Bunyaviridae genome structure and gene expression. *Curr Top Microbiol Immunol* 169:91–141
32. Fauquet CM, Fargette D (2005) International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Viol J* 2:64
33. Barr JN et al (2003) Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication. *Virology* 311(2):326–338
34. Barr JN, Wertz GW (2004) Bunyamwera bunyavirus RNA synthesis requires cooperation of 3'- and 5'-terminal sequences. *J Virol* 78(3):1129–1138
35. Kohl A et al (2006) Genetic elements regulating packaging of the Bunyamwera orthobunyavirus genome. *J Gen Virol* 87(Pt 1):177–187
36. Akashi H et al (1997) Sequence determination and phylogenetic analysis of the Akabane bunyavirus S RNA genome segment. *J Gen Virol* 78(Pt 11):2847–2851
37. Matsumori Y et al (2002) Serological and genetic characterization of newly isolated Peaton virus in Japan. Brief report. *Arch Virol* 147(2):401–410
38. Lee E et al (2000) Mutagenesis of the signal sequence of yellow fever virus prM protein: enhancement of signalase cleavage In vitro is lethal for virus production. *J Virol* 74(1):24–32
39. Botha EM et al (2008) Genetic determinants of virulence in pathogenic lineage 2 west nile virus strains. *Emerg Infect Dis* 14(2):222–230

Supplementary Material

Fig. S1 Amino acid alignment of the M glycoprotein protein precursor of members of the Simbu serogroups, indicating the gene layout and functional regions

*

10 20 30 40 50 60 70 80 90 100
SHUV (KF153117) ---MFLRVPIIPALNTGVPIPLKGGTTGSRFCFANGELVKTNTVTLIPSEVCKDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SHUV (HE800142) ---MFLRVPIIPALNTGVPIPLKGGTTGSRFCFANGELVKTNTVTLIPSEVCKDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
AINOV (HE795088) ---MFFKVLIFLTKCQKQKQKPLRGTTGSRFCFANGELVKTNTVNAIVPAEVCVDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SATV (HE795103) ---MLLNIILSNLAYLALFTPLKGGTGRSFCFANGELVKTNTVTSKVSCECCVDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SBV (HE649913) ---MLLNIILSNLAYLALFTPLKGGTGRSFCFANGELVKTNTVTSKVSCECCVDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
DOUV (HE795091) ---MLLNIILSNLAYLALFTPLKGGTGRSFCFANGELVKTNTVTSKVSCECCVDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SHAV (HE795106) MKTILRIASLTAQCAIMLCLPLRN---SIGORCFQGGPEFTINATSAPESEVCKDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SINV (HE795109) ---MPSVYVQTLIICATMIVLPLKGGTGRSFCFANGELVKTNTVTSKVSCECCVDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SHV (HE795100) ---MNFPLLIFLIPALVSPFEN---TKGGCFYGGKFKKVNATMNAEAVCLRDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
PEAV (HE795094) ---MNFPLLIFLIPALVSPFEN---TKGGCFYGGKFKKVNATMNAEAVCLRDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
SABOV (HE795097) ---MFLQIAILFVGLGSGMSPARN---NGAKCFQGGPEFTINATSAPESEVCKDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
JATV (QJ675602) ---MFLQIAILFVGLGSGMSPARN---NGAKCFQGGPEFTINATSAPESEVCKDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
OROV (NC005775) ---MFLQIAILFVGLGSGMSPARN---NGAKCFQGGPEFTINATSAPESEVCKDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
TINV (AB208700) ---MNTLILLLIYLVSSVMSPARN---NGGKCFYGGKFKKVNATMNAEAVCLRDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG
AKAV (NC009895) ---MIIITLNLVTLVVVTAMPFEN---NGGKCFYGGKFKKVNATMNAEAVCLRDDISLKSNGEHWKDDI---LGAIVKYRYLVQIKDWATCNPLIDNHG

510 520 530 540 550 560 570 580 590 600
SHUV (KF153117) IEKVGQCIATHQNTBAKMLLEDLILNYSIVDQGEKAEIKNANIN---CKTINKALBQSVLETAQAFYEQVSKMCPFDVMDITKPNASHLQWKTARTY
SHUV (HE800142) IEKVGQCIATHQNTBAKMLLEDLILNYSIVDQGEKAEIKNANIN---CKTINKALBQSVLETAQAFYEQVSKMCPFDVMDITKPNASHLQWKTARTY
AINOV (HE795088) IQLKISQCIATHQNTBAKMLLEDLILNYSIVDQGEKAEIKNANIN---CKTINKALBQSVLETAQAFYEQVSKMCPFDVMDITKPNASHLQWKTARTY
SATV (HE795103) IDHLKSCMAFYQNTKNSPVAKSLEVILOQMSVQQQEKBEIMSEIR---CTNINKITBLSVLETAQAFYEQVSKMCPFDVMDITKPNASHLQWKTARTY
DOUV (HE649913) IDHLKSCMAFYQNTKNSPVAKSLEVILOQMSVQQQEKBEIMSEIR---CTNINKITBLSVLETAQAFYEQVSKMCPFDVMDITKPNASHLQWKTARTY
SHAV (HE795106) IEBASACISYQNTQKQKLEIKSMEVLSDEVLILPQVPEVS---YINLILHEIENHDLVTAKEIILANLYEIVKVKYSAAGIQTWKTARTY
SINV (HE795109) IEQISKCLAVYQNTKSDRKEQVYLSLQYATK---BEISLNIIPDDELENTKIQEVLTKGKFLKLEYSKQTLSEBVDNKLINDYPAQKIKWYIKHR
SHV (HE795100) IEBSLCLALYQNTQNTYEQSLMTKLKSHDEBNLMLPTTKPT---FKEMIKKIEIANDVHMSITVEVNGMFPYETLKLKPAQETIEWRTYFQNT
PEAV (HE795094) IEBSLCLALYQNTQNTYEQSLMTKLKSHDEBNLMLPTTKPT---FKEMIKKIEIANDVHMSITVEVNGMFPYETLKLKPAQETIEWRTYFQNT
SABOV (HE795097) IEASGLICGVYKINRAKETSINLESEIENSHVETALMEQQET---FDLNLNKAFLDKLDTLAEVLYASVLEKQMSKHLDPAGENSIGQRAFLQNH
JATV (QJ675602) IEQISKCLAVYQNTKSDRKEQVYLSLQYATK---BEISLNIIPDDELENTKIQEVLTKGKFLKLEYSKQTLSEBVDNKLINDYPAQKIKWYIKHR
OROV (NC005775) IYEGLHNCIGPKMGIDYKQSKSELVDFSKNVLVDEYDKYFPE---DVTMNDIKPEIQAQDARMLLEERLYAKTE---CEMLLKKNSGQVFNMAVARTYLNKH
TINV (AB208700) VEDASLIGYQNTQNTYEQSLMTKLKSHDEBNLMLPTTKPT---FKEMIKKIEIANDVHMSITVEVNGMFPYETLKLKPAQETIEWRTYFQNT
AKAV (NC009895) IDDAJMCIGYQNTQNTYEQSLMTKLKSHDEBNLMLPTTKPT---FKEMIKKIEIANDVHMSITVEVNGMFPYETLKLKPAQETIEWRTYFQNT

110 120 130 140 150 160 170 180 190 200
SHUV (KF153117) TFMLLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
SHUV (HE800142) TFMLLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
AINOV (HE795088) TFMLLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
SATV (HE795103) SFMILLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
SBV (HE649913) SFMILLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
DOUV (HE795091) SFMIVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
SHAV (HE795106) TFLVILNIEDSGSISNACRYACRCEITTEQDNTGTVLISEEHLNYSIQOTTIKMGWFGKAIIEVLSLEATECHIHVTCQHTVEFNACRTHKKIRYFKGS
SINV (HE795109) TFLVMDVHVSIVAKRYACRCEITTEQDNTGTVLISEEHLNYSIQOTTIKMGWFGKAIIEVLSLEATECHIHVTCQHTVEFNACRTHKKIRYFKGS
SHV (HE795100) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
PEAV (HE795094) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
SABOV (HE795097) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
JATV (QJ675602) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
OROV (NC005775) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
TINV (AB208700) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS
AKAV (NC009895) TFMVLDINTGMLPKMHTCRCECDITLNDKTAIEILNSYRLNHRVSIOSMHISGWFKNKIDPLENCTCETIDVTCGLKTMSPHACFTHKSCIRYFKGS

610 620 630 640 650 660 670 680 690 700
SHUV (KF153117) TIALCNEHPHKHCKMCAEYTCYTSNTDHDGMRKVEYKHDINFEHDVIVLIRIKVMVPLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
SHUV (HE800142) TIALCNEHPHKHCKMCAEYTCYTSNTDHDGMRKVEYKHDINFEHDVIVLIRIKVMVPLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
AINOV (HE795088) TGLVCNQHPHKHCKMCAEYTCYTSNTDHDGMRKVEYKHDINFEHDVIVLIRIKVMVPLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
SATV (HE795103) TLGVCNQHPHKHCKMCAEYTCYTSNTDHDGMRKVEYKHDINFEHDVIVLIRIKVMVPLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
DOUV (HE649913) TLGVCNQHPHKHCKMCAEYTCYTSNTDHDGMRKVEYKHDINFEHDVIVLIRIKVMVPLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
SHAV (HE795106) GHLNCSHVHVKICRALLQBCQCSVTDSDGQLETVYKSHKEEYEDMASIKYVAFAPGLTKLVLNRSMSKALQAVDGLKLYTVANNHHLGI
SINV (HE795109) LNLNCPHSYKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
SHV (HE795100) GHLNCSHVHVKICRALLQBCQCSVTDSDGQLETVYKSHKEEYEDMASIKYVAFAPGLTKLVLNRSMSKALQAVDGLKLYTVANNHHLGI
PEAV (HE795094) SHLICQKIHKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
SABOV (HE795097) SHLICQKIHKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
JATV (QJ675602) SFLICQKIHKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
OROV (NC005775) SFLICQKIHKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
TINV (AB208700) SFLICQKIHKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV
AKAV (NC009895) SFLICQKIHKICRCHISPKYCSSTAHTHESELTIVYRDAEYTHLNLILDAKIVAPGLGSLTLQKIEBSRYSELVHIVKLLPKAKHIVQKOV

210 220 230 240 250 260 270 280 290 300
SHUV (KF153117) MLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SHUV (HE800142) MLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
AINOV (HE795088) MLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SATV (HE795103) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SBV (HE649913) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
DOUV (HE795091) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SHAV (HE795106) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SINV (HE795109) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SHV (HE795100) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
PEAV (HE795094) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
SABOV (HE795097) YMPYAMVBLNSNTELIIPVSIPLCLAFPIITRYIAYLLIPFVPIYVYQYLVKLCRTCPNIIALLHPTNCKICGSRFSNTEALVHRQCN
JATV (QJ675602) YMPYAMVBLNSNTELIIPVSIPLCLAFPIITRYIAYLLIPFVPIYVYQYLVKLCRTCPNIIALLHPTNCKICGSRFSNTEALVHRQCN
OROV (NC005775) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
TINV (AB208700) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN
AKAV (NC009895) VLPEITIBMCONTELIIGAVIFPQSIPLITLTVYVLPFIPYVVKYALVYRNFVFLCKKCLLVAHPFTNCPSTICGMVNTSBEALVHRQCN

710 720 730 740 750 760 770 780 790 800
SHUV (KF153117) LQFATQTLVWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
SHUV (HE800142) LQFATQTLVWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
AINOV (HE795088) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
SATV (HE795103) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
DOUV (HE649913) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
SHAV (HE795106) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
SINV (HE795109) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
SHV (HE795100) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
PEAV (HE795094) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
SABOV (HE795097) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
JATV (QJ675602) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
OROV (NC005775) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
TINV (AB208700) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH
AKAV (NC009895) LPEFNDHVLWYS---IVSEITNIVAMSLKESQGVSTQKLPOTAPLNTICSHKVICFSPRISQIYDIECKDLYL---KMFQDGVYRNHNSGBCACARDH

310 320 330 340 350 360 370 380 390 400
SHUV (KF153117) SCDQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SHUV (HE800142) SCDQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
AINOV (HE795088) TCDQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SATV (HE795103) NCSQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SBV (HE649913) NCSQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
DOUV (HE795091) NCSQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SHAV (HE795106) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SINV (HE795109) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SHV (HE795100) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
PEAV (HE795094) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
SABOV (HE795097) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
JATV (QJ675602) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
OROV (NC005775) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
TINV (AB208700) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI
AKAV (NC009895) NCTQYKALPKARRLCKSRVSNIIICACSAIIFPPIPTIN---ABCPTLBDLDEPFTACRQGVBEED---RMSKLRVVDDIITLILGCMTPPMTKLIKAI

810 820 830 840 850 860 870 880 890 900
SHUV (KF153117) CISTEFPAT---RQVRRRCESVETTYADDYSNAISEBCITVKTGTC---VKSSTWFPAVQC---GYVYTSARQHSKTHDITKCYLSTCQKRRVPPRSDVCS
SHUV (HE800142) CISTEFPAT---RQVRRRCESVETTYADDYSNAISEBCITVKTGTC---VKSSTWFPAVQC---GYVYTSARQHSKTHDITKCYLSTCQKRRVPPRSDVCS
AINOV (HE795088) CISTEFPAT---RQVRRRCESVETTYADDYSNAISEBCITVKTGTC---VKSSTWFPAVQC---GYVYTSARQHSKTHDITKCYLSTCQKRRVPPRSDVCS
SATV (HE795103) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
DOUV (HE649913) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
SHAV (HE795106) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
SINV (HE795109) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
SHV (HE795100) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
PEAV (HE795094) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
SABOV (HE795097) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
JATV (QJ675602) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
OROV (NC005775) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
TINV (AB208700) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS
AKAV (NC009895) CLEMFEPAE---RTISTTKICKIDIAVSESEYSTVPSNCKVRPGSC---VRHQGWIAECSNGLVYYSAKHNSKNDITLYLCSANCLDRVAFRSTSCS

410 420 430 440 450 460 470 480 490 500
SHUV (KF153117) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVDAND---KVDCTNLKDDQDIA
SHUV (HE800142) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
AINOV (HE795088) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
SATV (HE795103) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
SBV (HE649913) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
DOUV (HE795091) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
SHAV (HE795106) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
SINV (HE795109) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
SHV (HE795100) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
PEAV (HE795094) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
SABOV (HE795097) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
JATV (QJ675602) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
OROV (NC005775) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
TINV (AB208700) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA
AKAV (NC009895) VVVCFCGLMKRRRLKMGADGNTYCLMCICTDRGL-----TYHRASTRCYSPTKIKIYISWLLMLLNTALITVAEN---KVDCTNLKDDQDIA

910 920 930 940 950 960 970 980 990 1000
SHUV (KF153117) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
SHUV (HE800142) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
AINOV (HE795088) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
SATV (HE795103) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
DOUV (HE649913) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
SHAV (HE795106) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
SINV (HE795109) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
SHV (HE795100) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
PEAV (HE795094) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
SABOV (HE795097) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
JATV (QJ675602) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
OROV (NC005775) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
TINV (AB208700) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES
AKAV (NC009895) NTVVDSYTRKLNKHIHSHEDENYLSAQLSDIANDLTHHFRPTKMLPSADPTNIOITQDKISGSRINAIIEGKLPALAIAGLSDLVHMDPDTNLES

```

1100      1120      1130      1140      1150      1160      1170      1180      1190
SHUV (KF153117) I I I Y V K K I S I K S N Y Q I Y S T G P T V S I N V K H N E C T G C K P E S I P S D - L N W L T F S R E H S S S W G C E W G C I A I N E G C V F G S C Q D V I K P E M K V Y K R I G S E T K E V
SHUV (HE800142) I I I Y V K K I S I K S N Y Q I Y S T G P T V S I N V K H N E C T G C K P E S I P S D - L N W L T F S R E H S S S W G C E W G C I A I N E G C V F G S C Q D V I K P E M K V Y K R I G S E T K E V
AINOV (HE795088) L V V Y I K K I T V K S N Y Q I Y S T G P T I S V N V K H N E G C T G K C P T S I P A D - T N W L T F S R E R S S T W G C E W G C M A I N D G C V Y G S C Q D I I R P E L K I Y K K I G S E T K E A
SATV (HE795103) I V I V Y K V I N K A S Y T F L Y E T G P T I G I N A K H E B V C T G K C P T I P H Q - D G W I T F S K E R S S N W G C E W G C L A I N D G C L Y G S C Q D V I R P E Y K I Y K K S G I E Q K D V
SBV (HE649913) I I I V Y R R V I N K A S Y R F L Y E T G P T I G I N A K H E B V C T G K C P S I P H Q - D G W V T F S K E R S S N W G C E W G C L A I N D G C L Y G S C Q D I I R P E Y K I Y K K S I E Q K D V
DOUV (HE795091) I I I V Y R V I N K A S Y K F L Y E T G P T I G I N A K H E B V C T G K C P N I P H Q - D G W I T F S R E R S S N W G C E W G C L A I N D G C L Y G S C Q D I I R P E Y K I Y K K S G V E Q K D V
SHAV (HE795106) L V V Y I K R A T I K A E Y T F E Y E T G P T I G I N A V H S E K C T G C F R R I P H A - T N W L T F T K E H T S S W G C E W G C F A I N T G C V Y G S C Q D I I K P E G K V Y K K I G S E A I D A
S I M V (HE795109) L V V F V K K I I V K A Y T R H I Y T T G P T V G I N T K H N E I C T G C F K I P G E - K N W L T F S K E H T S N W G C E W G C L A I G A G C L Y G S C Q D I L K P E L R V Y K V G T E T R E V
SANV (HE795100) I V V Y V K Q A K I K A Q Y V Y Q S T G P T I G I N V K H E B L C T G S C P H D I P K D - Q N W L T F S K E H T S A W G C E W G C L A I G T G C V Y G S C Q D I I R E B G H V Y K K L G S E E I E T
PEAV (HE795094) I V V Y V K Q A K V K A Q Y V Y Q S T G P T I G I N V K H E B L C T G S C P H D I P K D - Q N W L T F S R E H T S A W G C E W G C L A I G T G C V Y G S C Q D I I K E B G H V Y K K L G S E E I E T
SABOV (HE795097) T V V Y I R L A Q V K A S Y A K E Y T G P T I A I N I H N E K C T G N C P E I I P K G E T W L S F S K E H T S A W G C E W G C L A V G T G C V Y G S C Q D V I R P E A T V Y K K S N Q E Q P D V
JATV (JQ675602) V I L F I R K A S Y R A E Y R R I Y T T G P T I G I N V E H N E C T G S C P K I E K K - Q N W L T F S K E H T S Q W G C E B Y G C L A M D A G C V Y G S C Q D I I R P E I D I Y K I Y G S E R S M V
OROV (NC005775) I V I V F K K A V K A Q Y K A Y T G P S I S I N I E H N E R C T G H C P E K I P A K - E G W L T F S K E H T S S W G C E B Y G C L A I D T G C L Y G S C Q D V I R P E L D V Y K K I G S E V S I D I
T I N V (AB208700) A I V I Y K S A K L K A T Y S L E Y Q G T P T I S I N T D H N E K C T G C P T V I P K K - D N W L T F S K E H T S A W G C E W G C I A I G T G C V Y G S C Q D V I R P E A K V Y K K I N Q D Q P I I
AKAV (NC009895) I I A Y V K S A R V I A T Y N H E Y K T G P T V S I N V Q H N E C T G C S P S I P K K - D N W L T F S R E H T S T W G C E W G C L A I G T G C V Y G S C Q D V I R E A T V I S R V N N E Q L E V

```

▼ Predicted signal peptide

```

1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
SHUV (KF153117) E V C I T T A H E T F C N N D V L Q P L I S Q R I Q D L L Q S I T T T M P P I I A V N K G K I Y T G D I N D L G V T A K K C G S V Q S T D N G I L G S G N V K F D Y I C H A F N R K D I I V R R C Y
SHUV (HE800142) E V C I T T A H E T F C N N D V L Q P L I S Q R I Q D L L Q S I T T T M P P I I A V N K G K I Y T G D I N D L G V T A K K C G S V Q S T D S G I L G S G N V K F D Y I C H A F N R K D I I V R R C Y
AINOV (HE795088) E V C I T T A H E T F C N T D V L Q P L I S Q R I Q D L Q T V T T A M M P N I I G V K D G K I Y S G D I N D L G T A K K C G S V Q M T D N S I I G T G N V K F D Y I C H A F N R K D I I V R K C F
SATV (HE795103) E V C I T M A H E S F C S T V D V L Q P L I S D R I Q L D I Q T V Q M S M P N I I A V N K G K I Y V G D I N D L G S T A K K C G S V Q L Y S E G V I G S G N K F D Y I C H A F N R K D V I L R R C F
SBV (HE649913) E V C I T M A H E S F C S T V D V L Q P L I S D R I Q L D I Q T I Q M D S M P N I I A V N K G V I Y V G D I N D L G S T A K K C G S V Q L Y S E G I I G S G T P K F D V I C H A F N R K D V I L R R C F
DOUV (HE795091) E V C I T M A H E S F C S T V D V L Q P L I S D R I Q L D I Q T V Q M E S M P N I I A V N K G K I Y V G D I N D L G S T A K K C G S V Q L Y S E G V I G S G N K F D Y I C H A F N R K D V I L R R C F
SHAV (HE795106) E I C I T D S S E T F C T E I T S Y N P I L E G K V Q I E I M S Q D S S L L P S N I F Q K N N N I Y K G D I N P K G T F A K K C G S V Q K V Q D Q I Y G S G E P R F D Y I C H A A S R R K D V V V R K C Y
S I M V (HE795109) E V C I T M P H E T F C N L D V L Q P L I S E R V Q I S L E T V D V K L P T L I G I N K G K V L G D I N D L G N T A K K C G S V Q I T S D S V I G S G T P K F D Y I C H A A Q R K D V I V R R C Y
SANV (HE795100) E I C I S D P Y E T L C N K I D V L E P I L T D T M Q V E I T S Q Q N I L L K E I V F V K D K N I Y S G D I N P K G T Y A K K C G S V Q K V G G L I S G I G N K F D V I C H A M S R R K D I I L R R C Y
PEAV (HE795094) E I C I S D P Y E T L C N K I D V L E P V I L T D T M Q V E I T A Q Q N I L L K E V L F V K D K N I Y S G D I N P R G T Y A K K C G S V Q K V D G L I S G I G N K F D V I C H A M S R R K D I I L R R C Y
SABOV (HE795097) E L C V T D P T G T L C N K I D A L E P I I G E H F Q V E L T T V Q T N F L P D V I L E K N K N I Y V G S I N R K G T S Q C C G S V Q Q G G D T Y V G N K F D V I C H A M S R R K D I V V R K C Y
JATV (JQ675602) E I C I T L P H E T F C N D L D L E P I I G D K I Q A A F Q T V Q S T H M P R L I A I K N N K A F T Q G I N D L G N T A S Y C S V Q K F N G T I L G Q D P K F D Y I C H A L R R K D I I V R R C Y
OROV (NC005775) E I C I T L P H E T F C N D M D L E P I I G D K L S A S F Q N T Q N G L P T L I A Y K K G K I Y T G Q I N D I G N T A L Q C G S I Q V I N G S T I G T G S P K F D Y I C H A M R R K D V I V R K C F
T I N V (AB208700) E L C V T D P S T L C N T I D A L E P I I G E H F Q V E I H T T Q N I L L P D R I L E K N K I Y T G Q I N A K G T F N P G C G S V Q L F D S N T Y I G N K F D Y I C H A L S R R K D I V V R K C Y
AKAV (NC009895) E F C V S E P T S T M C N T I N V L E P V L G E H M Q F E V H S V Q N I L L P E V A L I K N R R V Y K G S I N K K G V F N P Q C G S V Q S F D G K L Y G I G N P K F D Y I C H A L S R R K D I V V R K C Y

```

↳ Protein cleavage sites

■ N Glycosylation sites

— Trans-membrane regions

```

1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
SHUV (KF153117) E N A Y E S C K F L D M R E D L A M S G - M S N E V H M K I S N V G T I S Y K I M L G D F D Y D L F T E H A S L N L D A L K C G C C K S C P D G M H C S F K A N T D K I V L C K I S N C V S P L N
SHUV (HE800142) E N A Y E S C K F L D M R E D L A M S G - T S N E V H M K I S N V G T I S Y K I M L G D F D Y D L F T E H A S L N L D A L K C G C C K S C P D G M H C S F K A N T D K I V L C K I S N C V S P L N
AINOV (HE795088) D N A Y E S C K V L D L R N D L M Q A T - T N N E I H M K V S N V G T I N V Y M L G D F D Y D M F K S A A L T V D V L K C G C C L S C P D G M H C A F K A T T D K A T L C K I S N V S C I S P L N
SATV (HE795103) D N S Y Q S C L L E Q D S T L I A S T - N H M E V H K K V S V G T I N V K I M L G D F D Y N A Y S T Q A T I T I D E V R C S G C Y C P E G M A C S L K I T T N S I G S C P I K S N C D T Y I K I
SBV (HE649913) D N S Y Q S C L L E Q D N T L T I A S T - S H M E V H K K V S V G T I N V K I M L G D F D Y N A Y S T Q A T V T I D E R C S G C Y C P E G M A C A L K L S T N T I G S C P I K S N C D T Y I K I
DOUV (HE795091) D N S Y Q S C L L E Q D S T L N I A S T - S H M E V H K K V S V G T I N V K I M L G D F D Y N A Y S T Q A V A I D E R C S G C Y C P E G M A C S L K I T T N T I G S C P I K S N C D A Y I K I
SHAV (HE795106) D N A Y I S C T S L E I Q M N D L I K D E S K W I R Q D T G L Y G S V K W L L G D L N Y K O D D T T T R I T A K A V C S G C T D C I D D V S C R V D V T S N H I A S C S V E S T C S P Y I N R
S I M V (HE795109) D N A Y D S C R F L Q D P N Y N I A Q G - T A T E L H L K A N I Q Q I D F K M L G D F D Y D V S R D S T V I N S I K A C G C S C P E E M A C A L N I Q A E S S T T C K V T S N C E L M D R
SANV (HE795100) D N E Y I S C T L Q S E P D L K I K D N K T I I L N D Q L L G S A A R I N L N G D I K V K S V S A A E I Q I S G P T C Q C Q I Q C F T D A C K L E I T S T G I K Q C E I T S N C V V Y T N
PEAV (HE795094) D N E Y V S C A L Q L E P D L K I T D N K R I I L N D Q N L G G A A R I N L N G D I K V K S I S A A E I Q I S G T C Q C V C Q F T D A C K L E I T S T G I K Q C E V T S N C V V T S N
SABOV (HE795097) E N H Y F A C K T L E R Q D L K F Q K T K D N Q H L T D N I M L G S A K V R H I L G D I Q F S Q M P S R A Q L S G H A I C Q C Q D C P D N D V C K V T L T A E S S Y C K L I A S N C I S Y I E N
JATV (JQ675602) E N H Y E S C K F T P R P D I I D R K N D I L Q A S I V M N L Q M N F K I M L G D I N N Q G D I K V K S I S A A E I Q I S G T C Q C V C Q F T D A C K L E I T S T G I K Q C E V T S N C V V T S N
OROV (NC005775) E N D Y Q S C T L E K R N D L I P Y R K Q D I V E S K T G S N M G Q T F K I E L G D I N N K I P T K S I D - L Q M S I G A C C I D C A E I S C S I N A E V A E F V C H C K T N C E D P I N N
T I N V (AB208700) E N H Y Y A C S T L E R T D L R V H S K D Q Q V T N E M I I G T A T V R H I L G D I Q K S M S P V A D L S I N A V C Q C P D C F T D V I C K L I T T S N T V F C K V E S N C I S Y T D N
AKAV (NC009895) E N H Y Y S C A T L K A E A V I K P N I T N S K T M L N D N A L L G S A S V K I M L G D L I Y Q Q T S V Q E K R I D H A T C G C G T D C F N D V A C K I S M T S N O V Y Q C P I V S S C D S Y I N N

```

```

1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
SHUV (KF153117) I I I D P Q Q N D Y S I K L E C S E L I A D E I N I C H A K K A R P L I K Q V P K I S L A S I D E S T Y I E Q H D R C S T W L C R V R D E G I N A I L S P I F C K L S Y T R M V T I T V L L A I
SHUV (HE800142) I I I D P Q Q N D Y S I K L E C S Q L I N D E I S I C G L K L K A R P S I K Q N P K I S L A S I D E S T Y I E Q H D R C S T W L C R V R D E G I N A I L S P I F C K L S Y T R M V T I T V L L A I
AINOV (HE795088) I I I D P Q Q S E Y S I K L E C S Q L I N D E I S I C G L K L K A R P S I K Q N P K I S L A S I D E S T Y I E Q H D R C S T W L C R V R D E G I N A I L S P I F C K L S Y T R M V T I T V L L A I
SATV (HE795103) I A V D P M Q S D Y S I K L N C P L A T D T I S I N V C S A S V S T K P S I S K N P K I V L N S L D E T S Y I E Q H D K K C S T W L C R V R E G I S A I P F Q G L S Y T W A I M A S P L A I
SBV (HE649913) I A V D P M Q S D Y S I K L N C P L A T D T I S I S V C S A S T T K P S I S K N P K I V L N S L D E T S Y I E Q H D K K C S T W L C R V R E G I S V I F Q P L F G N L S Y T W K L T I Y I I I S L
DOUV (HE795091) I L S L E G S Q Q L H L K F R C - N M E I A I F S I C G I K A E I R S I V K S H K I L D L A S L S Q T S Y I R E Y D K K C G T W L C R V N E G I B O I I G P I W K E F N I W L K Y G T V V V V I I L
SHAV (HE795106) I I L D P A V N D Y S L K L C K N K D I K I K V A V C K S A S E T I P I L K Q S S V K L D L T S L D E S A Y I Q E H D K K C G T W L C R V K E G L S I I P E P I F Q K I S Y W K I P I A V T S A I
S I M V (HE795109) I I L Q G V N T L S I K L K - L S E K F K V S I C K N T I D L K S V V B D H S V L D T S S D Q T S Y I K E Y D K K C S T W L C R V R D E G L G F I P K P F W S E I N W G K Y I L I I V G L C
SANV (HE795100) I L V Q E G L M T L N V A L K L C - L S E K F V S I C K N T I D L K S V V B D H S V L D T S S D Q T S Y I K E Y D K K C S T W L C R V R D E G L G F I P K P F W S E I N W G K Y I L I I V G L C
PEAV (HE795094) Q V T E G T G D I S L K F R C - S T M V I T D L D C S I P I N V K S E I I K T S Q K D L T S A D Q T S Y I K E F D K C N T W L C R A Y N E G I S F I L Q P I W G E F T I W G Y L L I V A I I L
SABOV (HE795097) I L I Q P S V N D Y I K L T C R S K S D K L I N I C K E F V L P L T I K S H N Q I D L S K L D E S M Y I K E E D L R C E T M L C R V K D E G I S F I T E G I F G G I S Y F W K T I I S L A T I
JATV (JQ675602) V I I V F S P Q I K N Y N I K V C K S A V E K I T A H I C G R D I D L G L T I K P N Q K I D L S Q L D E S M Y I R E E D L Q C G T W L C R V Q E G I D I I P K G L F S G L R W T I I L Y S I G V
OROV (NC005775) I Q I P D G S S E I N L K F R C - S S T K D I K V C N T P V S V K S E I V K D R K L D L T S A D Q T S Y I K E F D K K C G T W L C R A Y N E G V S F I L G P L W N E I W G K Y V L I I S G L
T I N V (AB208700) V Y I N E G T N D V N L K F R C - L K A E I K I S I C G K E I P V K S E I I K D T K K L D W A S A D Q T S Y I K E F D K K A C A T W L C R A Y N E G I F V L E P L W N E L S W G K Y V L L A L A I I
AKAV (NC009895)

```

```

1410      1420      1430      1440
SHUV (KF153117) I S L I I L L Y V L I P F C K R I K G I L E Y N E R V Y Q I E N K F K
SHUV (HE800142) I S L I I L L Y V L I P F C K R I K G I L E Y N E R V Y Q I E N K F K
AINOV (HE795088) L L I V F L L F I L V P F C K R I K G V L E Y N E R I Y Q I E N K S K
SATV (HE795103) F I L I L T L Y I L I P L C K R I K G L L E Y N E R I Y Q M E N K F K
SBV (HE649913) I M L I L F L Y I L I P L C K R I K G L L E Y N E R I Y Q M E N K F K
DOUV (HE795091) F T L I L L Y I L I P L C K R I K G L L E Y N E R I Y Q M E N K L K
SHAV (HE795106) F S V I L I V K V I N P L I K L I I I N T L K H N E Q M Y L L E S K Q R
S I M V (HE795109) I G L L A L Y I C F P V C K R I K G L L E Q N E R I Y Q M E T K F K
SANV (HE795100) L L I V I V A K I A P F V K Y V I S T L K Y N D Y I Y K I E N K M H
PEAV (HE795094) L L I I I A K L I A P L I K Y I I S T L K Y N D Y V Y K I E N K M H
SABOV (HE795097) I L L I F A K V L K P L A P A I I T L L K R N D E I Y R L E N K I K
JATV (JQ675602) G I L V I F Y Y I G L P L F K I L K D Y L K R N E L E Y L A E R K L K
OROV (NC005775) V I I V I L Y I L L P I G R L K A F L I R H E I E Y A M E Q K I K
T I N V (AB208700) I I I V L F A K I I K P L S I Y L I S L L K N E E I Y R L E N K I K
AKAV (NC009895) I S V L I L V K V I R P L A R Y V V A V L K E N D K V I Y K L E N K L K

```