Genomic characterization of Sebokele virus 1 (SEBV1) reveals a new candidate species among the genus Parechovirus.

Marie-Line Joffret, Christiane Bouchier, Marc Grandadam, Hervé Zeller, Corinne Maufrais, Hervé Bourhy, Philippe Desprès, Francis Delpeyroux, Laurent Dacheux

Title

Genomic characterization of Sebokele virus 1 (SEBV1) reveals a new candidate species among the genus *Parechovirus*

Running title

Molecular characterization of Sebokele virus 1

Contents category : Short communication

Authors and affiliations

Joffret ML[1,2], Bouchier C[3], Grandadam M[4], Zeller H[5], Maufrais C[6], Bourhy H[7], Despres P[8], Delpeyroux F[1,2], Dacheux L[7*].

1 Institut Pasteur, Unité de Biologie des virus entériques, Paris, France

2 INSERM, U994, Paris, France

3 Institut Pasteur, Plate-forme Génomique, Paris, France

4 Institut Pasteur, Centre National de Reference des arbovirus, Unité des Interactions Moléculaires Flavivirus-Hôtes, Paris, France

5 Institut Pasteur, Unité de Biologie des Infections Emergentes, Lyon, France

6 Institut Pasteur, Centre d'Informatique pour les Biologistes, Paris, France

7 Institut Pasteur, Unité Dynamique des Lyssavirus et Adaptation à l'Hôte, Paris, France.

8 Institut Pasteur, Unité des Interactions Moléculaires Flavivirus-Hôtes, Paris, France

*** Corresponding author**

Email : laurent.dacheux@pasteur.fr

Phone : +33 1 40 61 33 03

Fax : +33 1 40 61 30 20

Word count: summary = 125, main text (with figure legends) = 2422

29

30 **Number of tables and figures** : 1 table (and 1 supplementary table), 2 figures (and 1 supplementary

31 figure)

32

33 **Footnote** : The EMBL accession number for the genome sequence of Sebokele virus 1 is HF677705.

34

35 **Keywords** : Sebokele virus 1, picornavirus, parechovirus, Ljungan virus, high throughput sequencing,

36 Illumina sequencing, complete genome, phylogenetic analysis

37

Summary

We determined the genomic features and the taxonomic classification of Sebokele virus 1 (SEBV1), a previously unclassified arbovirus isolated in 1972 from rodents collected in Botambi, Central African Republic. The complete genome sequence was obtained using a deep sequencing approach (Illumina technology) and dedicated bioinformatics workflows for data analysis. Molecular analysis identified SEBV1 as a picornavirus, most closely related to Ljungan viruses of the genus *Parechovirus*. The genome has a typical Ljungan virus-like organization, including the presence of two unrelated 2A protein motifs. Phylogenetic analysis confirmed that SEBV1 belongs to the parechovirus phylogroup and was most closely related to the *Ljungan virus* species. However, it appeared clearly distinct from all members of this phylogroup, suggesting that it represents a novel species of the genus *Parechovirus*.

Main text

The family *Picornaviridae* encompasses small nonenveloped and positive single-stranded RNA viruses, and includes many human and animal pathogens. The typical genome organization of picornaviruses includes a single open reading frame (ORF) encoding a large protein precursor (polyprotein), preceded by a 5' untranslated region (5'UTR) and followed by a 3'UTR and a polyadenylated tract of variable length (Knowles *et al.*, 2012; Racaniello, 2007). The polyprotein contains, in the following order, a non structural leader (L) protein in some cases, then a first domain P1 encoding three structural proteins (VP0, which is autocatalytically cleaved further into VP4 and VP2 in most picornaviruses, VP3 and VP1), and two distinct non structural coding regions (P2 encoding proteins 2A, 2B and $2C^{ATPase}$, a multifunctional ATPase, and P3 encoding proteins 3A, $3B^{VPg}$, $3C^{pro}$, a cysteine protease, and $3D^{pol}$, a RNA-dependent RNA polymerase). The family *Picornaviridae* is currently divided into 12 genera: *Aphtovirus*, *Avihepatovirus*, *Cardiovirus*, *Enterovirus*, *Erbovirus*, *Hepatovirus*, *Kobuvirus*, *Parechovirus*, *Sapelovirus*, *Senecavirus*, *Teschovirus* and *Tremovirus*. However, the diversity of this viral family is far from being fully documented, and numerous novel picornavirus species and genera have recently been described (Knowles *et al.*, 2012) (http://www.picornaviridae.com/).

The genus *Parechovirus* includes two species, *human parechovirus* (HPeV) and *Ljungan virus* (LV), each including various types and/or genotypes (Johansson *et al.*, 2002; Joki-Korpela & Hyypia, 2001; Stanway & Hyypia, 1999). HPeV is a pathogen infecting mainly the gastrointestinal and the respiratory tracts of children (Baumgarte *et al.*, 2008; Harvala & Simmonds, 2009; Harvala *et al.*, 2008; Joki-Korpela & Hyypia, 2001). Sixteen different types of HPeV and several different genotypes have been identified (Knowles *et al.*, 2012) (http://www.picornaviridae.com/). The genome organization of HPeV is typical of the family, except that the VP0 protein appears not to be cleaved and its N-terminus not myristoylated (Knowles *et al.*, 2012; Stanway & Hyypia, 1999).

Ljungan virus (LV) was first isolated from bank voles (*Myodes glareolus*) trapped in Sweden (Niklasson *et al.*, 1998; Niklasson *et al.*, 1999). Few isolates of the LV species have been described and only five whole genome sequences are available. They include three Swedish LV: strains 87-012 (the prototype strain) and 174F, both representing genotype 1, and 145SL representing genotype 2

(Johansson *et al.*, 2002). Two American LV isolates (M1146 and 64-7855) have also been sequenced (Johansson *et al.*, 2003; Tolf *et al.*, 2009; Whitney *et al.*, 1970), and represent distinct genotypes. Molecular characterization of LV isolates revealed genomic features unusual in picornaviruses: they have only three different structural proteins like all parechoviruses (Ekstrom *et al.*, 2007; Johansson *et al.*, 2004; Tolf *et al.*, 2008) and two different 2A protein motifs (Johansson *et al.*, 2003; Johansson *et al.*, 2002). LV has been suggested to be the aetiological agent of myocarditis and possibly other human diseases (Niklasson *et al.*, 1998; Niklasson *et al.*, 2007).

Sebokele virus 1 (SEBV1) was originally isolated in 1972 from the crushed organs of *Hylomyscus* sp. (African wood mice) collected in Botambi, Central African Republic (Digoutte, 1985; El Mekki *et al.*, 1981; Zeller *et al.*, 1989). The virus was isolated by intracerebral inoculation of suckling mice (Digoutte, 1978). The pathogenic agent appeared to be resistant to chloroform and could be filtered through 220 nm-pore size filters. Attempts to identify the virus using sera against arboviruses and herpesvirus, and by electron microscopy after inoculation of cell cultures, were unsuccessful. The isolate was therefore described to be a non-identified arbovirus. Three similar viral isolates were reported from *Muridae* in the Central African Republic: two from *Hylomyscus* sp. and one from *Praomys* sp. (Digoutte, 1985).

We characterized the viral genome of SEBV1 by deep and classical Sanger sequencing approaches. Total RNA (TRI Reagent) was extracted from a vial, dated 1974, of a lyophilized 10% suspension of the original inoculated suckling mouse brain. The cDNA was synthesised and subjected to sequence-independent amplification as described previously (Dacheux *et al.*, 2010). Amplified DNA was then used for Illumina single-read sequencing (69 nt single reads, GAIIx). The Illumina Analysis Pipeline version 1.6 was used for image analysis, base calling, error estimation and demultiplexing.

A total of 8,509,420 reads were obtained, trimmed and mapped to the *Mus musculus* genome by using CLC Assembly Cell (v. 3.11) with default parameters. Non-mapped reads (147,826 reads) were extracted and assembled into contigs by three different assemblers (CLC novo assembler, VelvetOptimiser and SOAP de novo). The contigs were used independently for BLASTn and BLASTx identity searches against EMBL and UniProt databases, respectively. The blast2taxoclass program (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::blast2taxoclass) was used for taxonomic

106 classification, and demonstrated that most of the viral contigs were related to the family
107 *Picornaviridae*, strongly suggesting that SEBV1 belonged to this family. The blast2genoclass program
108 (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::blast2genoclass) was used for genomic annotation
109 of contigs and high-scoring segment pairs (hsps) matching with picornaviruses were extracted.
110 Extracted hsps were assembled using Sequencher 5.0 software (Gene Codes Corporation) with default
111 parameters, resulting in the generation of three main contigs covering nearly 99% of the predicted
112 genomic sequences (Fig 1a). Following BLASTx analysis, contigs presented a maximum amino acid
113 (aa) identity of 61 to 65% with parechoviruses.
114 Attempts to culture the virus in several cell lines were all unsuccessful, so a second vial of SEBV1
115 was used to finalise the full-length genome sequence determination. Total RNA was extracted,
116 amplified by RT-PCR with primers corresponding to contig sequences, and subjected to Sanger
117 sequencing (Supplementary Table 1). The final full-length sequence served as a reference for mapping
118 reads generated by deep sequencing (CLC Assembly Cell). The average coverage for each nucleotide
119 position was 104 times (lower at the extremities), without any discrepancies being detected (Fig. 1b).
120 SEBV1 genomic sequences were compared to reference nucleotide and protein sequences from the
121 different viral genera downloaded from GenBank. BioEdit software (Hall, 1999) was used to translate
122 DNA and the ClustalX program, version 2.0 (Larkin *et al.*, 2007; Thompson *et al.*, 1997) for sequence
123 alignments. Pairwise alignments were performed using CLC Main Workbench 5.7.2 software (CLC
124 bio). Phylogenetic analyses of aa sequences were conducted by the neighbor-joining algorithm using
125 MEGA 4.0.2 (Tamura *et al.*, 2007) with a Poisson correction model and 1,000 bootstrap replicates.
126 Trees were visualised using the FigTree program, version 1.3.1 (available from
127 http://tree.bio.ed.ac.uk/software/figtree). The GenBank accession numbers of the viral sequences used
128 in the phylogenetic analyses are indicated in the trees.
129 The genome of SEBV1 was 7,537 nucleotides (nt) long (excluding the poly(A) tail), and is organised
130 as follows: a 5'UTR sequence of at least 754 nt, followed by an ORF of 6,702 nt predicted to encode a
131 polyprotein of 2,233 aa, terminated by a 3'UTR of 81 nt and a polyadenylated tract. The exact length
132 of the 5'UTR sequence remains however to be confirmed by functional analysis, as the use of
133 lyophilised stored vials as the starting material did not allow us to assume that the 5' end of the

genome sequence was complete. More than 30% of the 5'UTR sequence (nucleotide position 400 to 735) presented nt identities with 5'UTR sequences of parechoviruses, according to BLASTn analysis, with maximum coverage (43%) and nt identity (72%) obtained with LV strain 64-7855. The 3'UTR sequence of SEBV1 seems to be unique, and BLASTn searches found no related sequence. The G+C content of the SEBV1 genome is 45.7%, higher than the G+C contents of LV (41.6-42.5%) and human parechoviruses (39.0-39.5%). BLASTp analysis of the complete deducted polyprotein identified the best matches as LV species (the best being LV strain 87-012), with 62-63% aa identity and 77-78% aa similarity, and human parechoviruses HPeV-1 and -6 exhibited nearly 46% aa identity and 63% aa similarity. The polyprotein sequence of SEBV1 was then aligned with its closest relatives: the Ljungan viruses (strains 87-012, 174F, 145SLG, M1146 and 64-7855) and the human parechoviruses HPeV-1 (strain Harris) and HPeV-6 (strain NII561-2000). With NetPicoRNA prediction (Blom *et al.*, 1996), this allowed the identification of potential cleavage sites and processing of this polyprotein (see Table 1 and Fig. 1c). The organisation of the SEBV1 genome is typical of picornaviruses, with a P1 region encoding the structural proteins and two non structural regions, P2 and P3.

The P1 region is the genomic region of SEBV1 most divergent from the most closely related picornaviruses, with nt identities of 56-57% with LV isolates and 36-37% with the most closely related HPeV (Table 1). This region encodes the classical capsid proteins with, in order, VP0, VP3 and VP1. The P1 region contains the picornavirus capsid protein domain-like (pfam sequence cluster cd00205 - Rhv-like) sequence found in other species of the genera *Parechovirus*, *Avihepatovirus* and *Aquamavirus*, and other unclassified picornaviruses including turdivirus. VP0 of SEBV1 is probably not cleaved into VP4 and VP2. Similarly to the other parechoviruses, the N-terminal end of this capsid protein does not contain the myristylation motif (GxxxS/T) found in most picornaviruses (Chow *et al.*, 1987). Like LV and pasivirus 1 (SPaV1), VP0 of SEBV1 has a short N-terminal extremity (Sauvage *et al.*, 2012). The VP3 of SEBV1 has a long N-terminal extremity enriched with basic residues, like those of parechoviruses (Johansson *et al.*, 2002). However, the highly conserved KxKxxRxK motif in the parechovirus P1 region was not found, and is replaced by the motif R_{270}FKWTRNN (Johansson *et al.*, 2002; Williams *et al.*, 2009). VP1 is predicted to be cleaved at the canonical cleavage site DxExNPG_{812}P (Fig. 1c) (Ryan & Flint, 1997). The parechovirus (PS)ALxAxETG motif was not

present in this capsid protein. As in LV, SPaV1 and some HPeVs, SEBV1 lacked the RGD tri-peptide involved in integrin receptor binding, raising questions about the mechanism of cellular entry and the host tropism of this virus (Knowles *et al.*, 2012; Merilahti *et al.*, 2012; Sauvage *et al.*, 2012; Williams *et al.*, 2009). The VP1 protein of SEBV1, like those of LV and SPaV1, contained one N-terminal insertion (8 aa long) and a unique C terminal extension (48 aa long). Similarly to LV, this protein contains the $3C^{pro}$ cleavage site $RRQ_{794}GCR$ predicted by NetPicoRNA at the end of its N-terminus, suggesting the existence of a short 2A-like protein (18 aa) (Fig. 1c).

The non structural P2 region of SEBV1 shows 70-72% aa sequence identity with LV and nearly 48% with HPeV. This region was deduced to encompass the non-structural proteins 2A, 2B and 2C (Table 1 and Fig 1c). The 2A protein possesses the highly conserved picornaviral H-box/NC regions, involved in the control of cell proliferation (Hughes & Stanway, 2000). The absence of the conserved GxCG motif suggests that the 2A protein of SEBV1 does not have trypsin-like proteolytic activity (Lamphear *et al.*, 1993). The 2C protein, like those of other parechoviruses, carries the SxxGxGKx NTase motif and the D_{1271}DLxQ helicase motif (Gorbalenya *et al.*, 1989a).

The P3 region encompasses the non-structural proteins 3A, 3B, 3C and 3D, displaying 60-62% amino acid identity with LV and nearly 42% with HPeV (Table 1 and Fig 1c). The protein 3B contains the conserved tyrosine (Y) residue at the third position of the predicted N-terminus, present in all known picornaviruses and necessary for the priming function of the VPg (Ambros & Baltimore, 1978). The conserved catalytic triad H_{1610}-D_{1648}-C_{1723} identifies the 3C protein as a cysteine-protease (Bazan & Fletterick, 1988) with the active site motif G_{1721}xCG and G_{1739}xH (Gorbalenya *et al.*, 1989b). The 3D protein contained the five highly conserved RNA polymerase motifs: K_{1930}DELR, G_{2057}GxPSG, $S_{2061}GX_3TX_3N$, Y_{2096}GDD and F_{2144}LKR (Koonin, 1991).

Phylogenetic analysis based on the most informative picornaviral protein sequence, *i.e.* VP1, of SEBV1 and representative members of the family *Picornaviridae* confirmed that this virus is related to the genus *Parechovirus* (Fig. 2). SEBV1 appears to be distinct from the two species constituting this genus, albeit more closely related to LV. It appears at the basal position of the LV phylogroup in the phylogenetic tree, suggesting an ancestral position. Phylogenetic analyses of $3D^{pol}$ sequences (data

not shown) and of the complete structural region P1 in representative picornaviral sequences gave similar results (Supplementary Fig. 1).

This study illustrates the value of deep sequencing approaches (Illumina technology) for the identification and the taxonomic classification of unclassified or misclassified viruses. This approach is useful when no genomic information is available and even when using the sample originally collected as the starting material. We developed and validated a workflow for data analysis, from raw read sequences to the generation of large contig sequences covering almost all of the genome. The aa identity between SEBV1 and LV isolates was over 55% in the polyprotein, and over 50% in P1, P2 or P3. SEBV1 exhibited the same genome organization as LV, with some common genetic features. Both SEBV1 and LV were originally isolated from rodents, suggesting that they share similar host tropism. These various findings indicate that SEBV1 is a member of the genus *Parechovirus*, closely related to *Ljungan virus* species

Despite these common features, the aa identity between SEBV1 and other known parechoviruses was below 70% for the polyprotein and in particular for the P1 protein. The G+C content of the SEBV genome differs from those of parechoviruses by more than 1%. Moreover, SEBV1 exhibits features not shared with those of LV and other parechoviruses. We therefore suggest that SEBV1 belongs to a new distinct species of parechovirus, in line with the criteria for species demarcation defined by the International Committee for Taxonomy of Viruses.

Identification of this putative third species in the genus *Parechovirus* highlights the viral and genetic diversity of this genus. It is interesting also to note that SEBV1 was the first LV-like virus to be isolated in Africa. However, further information is needed about this newly characterised picornavirus, especially regarding its host range, geographical distribution and genetic diversity.

Legends to the figures

Fig. 1. Schematic representation of predicted genome organisation of SEBV1. (a) The three large sequence contigs obtained after *de novo* assembly and BLAST analysis are indicated with black lines. The positions of primers designed to generate overlapping PCR products and for genome resequencing are shown above the sequence contigs. (b) Representation of the genome coverage of SEBV1 after Illumina sequencing, with sequence reads mapped on the full-length genome sequence. The number of times that each nucleotide position was sequenced is indicated on the ordinate. (c) Predicted genome organization of SEBV1. Positions of predicted cleavage sites are indicated along the polyprotein, in nucleotides (upper numbers) and amino acids (lower numbers). Conserved picornaviral amino acid motifs are presented below the bar.

Fig. 2. Phylogenetic analysis based on the SEBV1 VP1 protein (310 aa) with representative members of the genera of the family *Picornaviridae*, and also some unassigned picornaviruses. The scale bar indicates branch length, and bootstrap values ≥ 60% are shown next to the relevant nodes. The tree is midpoint rooted for purposes of clarity only.

Fig. S1. Phylogenetic analysis based on the P1 region (812 aa) of SEBV1 and representative members of selected genera in the family *Picornaviridae*, with some additional unassigned picornaviruses. The scale bar indicates branch length, and bootstrap values ≥ 70% are shown next to the relevant nodes. The tree is midpoint rooted for purposes of clarity only.

Acknowledgements

We are grateful to Laurence Ma (Plate-forme Génomique) for her excellent technical expertise and Nick J. Knowles for information about new parechovirus species.

We are grateful for the financial support of the *Agence Nationale pour la Recherche* (ANR 09 MIEN 019) and the *Fondation pour la Recherche Médicale* (FRM DMI20091117313).

References

Ambros, V. & Baltimore, D. (1978). Protein is linked to the 5' end of poliovirus RNA by a phosphodiester linkage to tyrosine. *The Journal of biological chemistry* **253**, 5263-5266.

Baumgarte, S., de Souza Luna, L. K., Grywna, K., Panning, M., Drexler, J. F., Karsten, C., Huppertz, H. I. & Drosten, C. (2008). Prevalence, types, and RNA concentrations of human parechoviruses, including a sixth parechovirus type, in stool samples from patients with acute enteritis. *Journal of clinical microbiology* **46**, 242-248.

Bazan, J. F. & Fletterick, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7872-7876.

Blom, N., Hansen, J., Blaas, D. & Brunak, S. (1996). Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein science : a publication of the Protein Society* **5**, 2203-2216.

Chow, M., Newman, J. F., Filman, D., Hogle, J. M., Rowlands, D. J. & Brown, F. (1987). Myristylation of picornavirus capsid protein VP4 and its structural significance. *Nature* **327**, 482-486.

Dacheux, L., Berthet, N., Dissard, G., Holmes, E. C., Delmas, O., Larrous, F., Guigon, G., Dickinson, P., Faye, O. & other authors (2010). Application of broad-spectrum resequencing microarray for genotyping rhabdoviruses. *Journal of virology* **84**, 9557-9574.

Digoutte, J. P. (1978). Sebokele (SEB). *The American journal of tropical medicine and hygiene* **17**, 424-425.

Digoutte, J. P., Germain, M (1985). Sebokele (An B 1227d), nouvel Arbovirus non groupé. In *International Catalogue of Arboviruses Including Certain Other Virus of Vertebrates*, Third edn, pp. 917-918. Edited by N. Karabatsos. San Antonio, Texas: Am. Soc. Trop. Med. Hyg.

Ekstrom, J. O., Tolf, C., Fahlgren, C., Johansson, E. S., Arbrandt, G., Niklasson, B., Edman, K. A. & Lindberg, A. M. (2007). Replication of Ljungan virus in cell culture: the genomic 5'-end, infectious cDNA clones and host cell response to viral infections. *Virus research* **130**, 129-139.

El Mekki, A. A., Nieuwenhuysen, P., van der Groen, G. & Pattyn, S. R. (1981). Characterization of some ungrouped viruses. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **75**, 799-806.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989a). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic acids research* **17**, 4713-4730.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989b). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS letters* **243**, 103-114.

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

Harvala, H. & Simmonds, P. (2009). Human parechoviruses: biology, epidemiology and clinical significance. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* **45**, 1-9.

Harvala, H., Robertson, I., McWilliam Leitch, E. C., Benschop, K., Wolthers, K. C., Templeton, K. & Simmonds, P. (2008). Epidemiology and clinical associations of human parechovirus respiratory infections. *Journal of clinical microbiology* **46**, 3446-3453.

Hughes, P. J. & Stanway, G. (2000). The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation. *The Journal of general virology* **81**, 201-207.

Johansson, E. S., Niklasson, B., Tesh, R. B., Shafren, D. R., Travassos da Rosa, A. P. & Lindberg, A. M. (2003). Molecular characterization of M1146, an American isolate of Ljungan virus (LV) reveals the presence of a new LV genotype. *The Journal of general virology* **84**, 837-844.

Johansson, E. S., Ekstrom, J. O., Shafren, D. R., Frisk, G., Hyypia, T., Edman, K. & Lindberg, A. M. (2004). Cell culture propagation and biochemical analysis of the Ljungan virus prototype strain. *Biochemical and biophysical research communications* **317**, 1023-1029.

Johansson, S., Niklasson, B., Maizel, J., Gorbalenya, A. E. & Lindberg, A. M. (2002). Molecular analysis of three Ljungan virus isolates reveals a new, close-to-root lineage of the Picornaviridae with a cluster of two unrelated 2A proteins. *Journal of virology* **76**, 8920-8930.

Joki-Korpela, P. & Hyypia, T. (2001). Parechoviruses, a novel group of human picornaviruses. *Annals of medicine* **33**, 466-471.

Knowles, N. J., Hovi, T., Hyypiä, T., King, A. M. Q., Lindberg, A. M., Pallansch, M. A., Palmenberg, A. C., Simmonds, P., Skern, T. & other authors (2012). Picornaviridae. In *Virus taxonomy: classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 855-880. Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. San Diego: Elsevier Academic Press.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *The Journal of general virology* **72 (Pt 9)**, 2197-2206.

Lamphear, B. J., Yan, R., Yang, F., Waters, D., Liebig, H. D., Klump, H., Kuechler, E., Skern, T. & Rhoads, R. E. (1993). Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of the 2A proteases from human Coxsackievirus and rhinovirus. *The Journal of biological chemistry* **268**, 19200-19203.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Merilahti, P., Koskinen, S., Heikkila, O., Karelehto, E. & Susi, P. (2012). Endocytosis of integrin-binding human picornaviruses. *Advances in virology* **2012**, 547530.

Niklasson, B., Hornfeldt, B. & Lundman, B. (1998). Could myocarditis, insulin-dependent diabetes mellitus, and Guillain-Barre syndrome be caused by one or more infectious agents carried by rodents? *Emerging infectious diseases* **4**, 187-193.

Niklasson, B., Samsioe, A., Papadogiannakis, N., Kawecki, A., Hornfeldt, B., Saade, G. R. & Klitz, W. (2007). Association of zoonotic Ljungan virus with intrauterine fetal deaths. *Birth defects research Part A, Clinical and molecular teratology* **79**, 488-493.

Niklasson, B., Kinnunen, L., Hornfeldt, B., Horling, J., Benemar, C., Hedlund, K. O., Matskova, L., Hyypia, T. & Winberg, G. (1999). A new picornavirus isolated from bank voles (Clethrionomys glareolus). *Virology* **255**, 86-93.

Racaniello, V. (2007). Picornaviridae: the viruses and their replication. In *Fields Virology*, Fifth edn, pp. 795-838. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott Williams and Wilkins.

Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *The Journal of general virology* **78 (Pt 4)**, 699-723.

Sauvage, V., Ar Gouilh, M., Cheval, J., Muth, E., Pariente, K., Burguiere, A., Caro, V., Manuguerra, J. C. & Eloit, M. (2012). A member of a new Picornaviridae genus is shed in pig feces. *Journal of virology* **86**, 10036-10046.

Stanway, G. & Hyypia, T. (1999). Parechoviruses. *Journal of virology* **73**, 5249-5254.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular biology and evolution* **24**, 1596-1599.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25**, 4876-4882.

Tolf, C., Gullberg, M., Johansson, E. S., Tesh, R. B., Andersson, B. & Lindberg, A. M. (2009). Molecular characterization of a novel Ljungan virus (Parechovirus; Picornaviridae) reveals a fourth genotype and indicates ancestral recombination. *The Journal of general virology* **90**, 843-853.

Tolf, C., Ekstrom, J. O., Gullberg, M., Arbrandt, G., Niklasson, B., Frisk, G., Liljeqvist, J. A., Edman, K. & Lindberg, A. M. (2008). Characterization of polyclonal antibodies against the capsid proteins of Ljungan virus. *Journal of virological methods* **150**, 34-40.

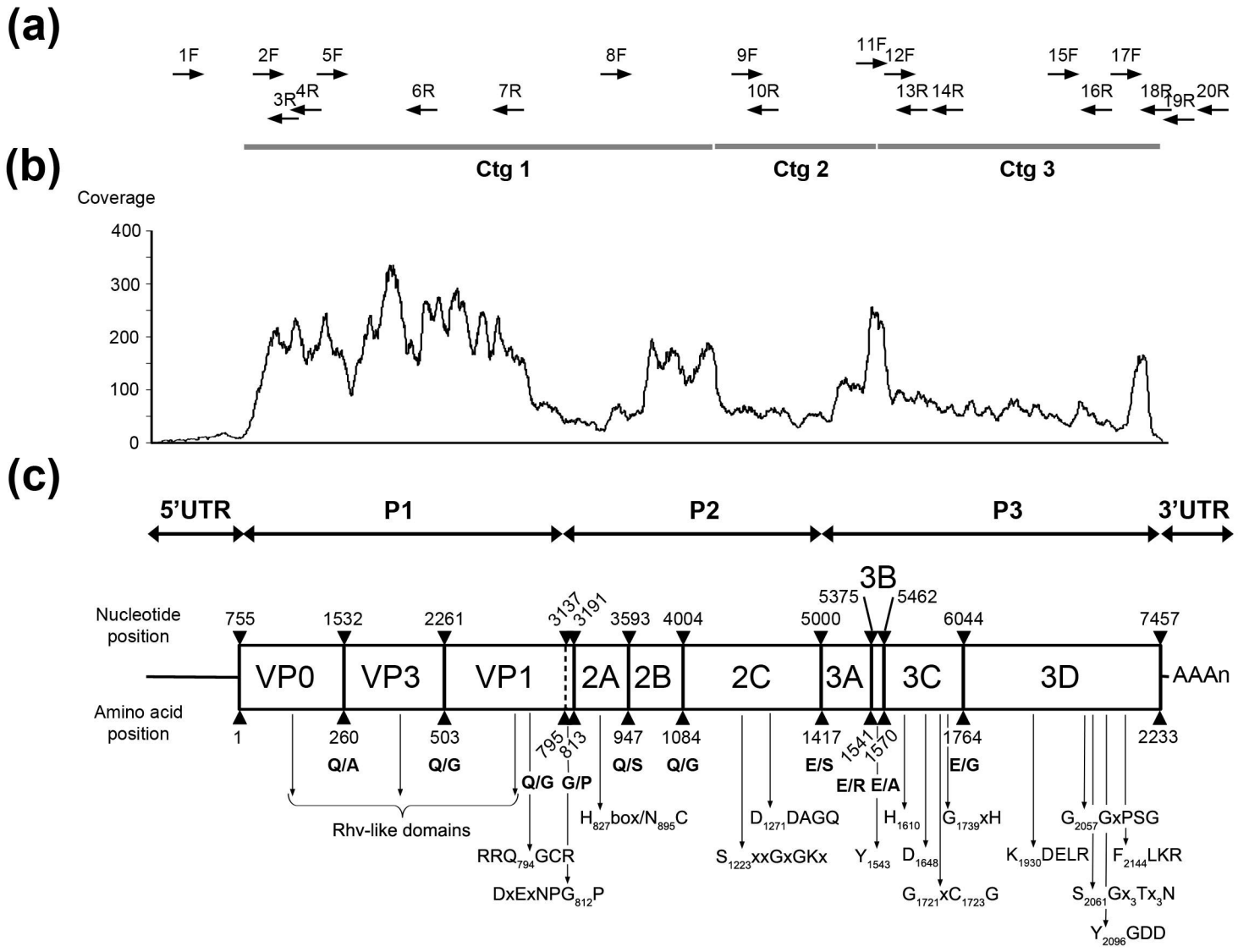
Whitney, E., Roz, A. P. & Rayner, G. A. (1970). Two viruses isolated from rodents (Clethrionomys gapperi and Microtus pennsvlvanicus) trapped in St. Lawrence County, New York. *Journal of wildlife diseases* **6**, 48-55.

Williams, C. H., Panayiotou, M., Girling, G. D., Peard, C. I., Oikarinen, S., Hyoty, H. & Stanway, G. (2009). Evolution and conservation in human parechovirus genomes. *The Journal of general virology* **90**, 1702-1712.

Zeller, H. G., Karabatsos, N., Calisher, C. H., Digoutte, J. P., Murphy, F. A. & Shope, R. E. (1989). Electron microscopy and antigenic studies of uncharacterized viruses. I. Evidence suggesting the placement of viruses in families Arenaviridae, Paramyxoviridae, or Poxviridae. *Archives of virology* **108**, 191-209.

Table 1. Pairwise amino acid identities between the predicted proteins of SEBV1 and related members of the genus *Parechovirus*.

Genomic features			**Pairwise amino acid identity (%) (length of protein in aa)**						
SBKV			**Related parechoviruses**						
Protein	**Position**	**Length**	**LV1** 87-012 (EF202833)	**LV1** 174F (AF327921)	**LV2** 145SLG (FJ384560)	**LV3** M1146 (AF538689)	**LV4** 64-7855 (EU854568)	**HPeV-1** Harris (L02971)	**HPeV-6** NII561-2000 (AB252582)
VP0	1 Met - 259 Gln	259	65.3 (259)	65.6 (259)	63.3 (259)	65.3 (259)	64.9 (259)	37.7 (289)	37.7 (289)
VP3	260 Ala - 502 Gln	243	61.1 (244)	61.5 (244)	62.3 (244)	60.7 (244)	62.3 (244)	48.8 (253)	48.0 (252)
VP1	503 Gly - 812 Gly	310	46.3 (317)	46.3 (317)	46.6 (317)	44.6 (319)	45.8 (319)	25.3 (234)	26.6 (234)
P1	**1 Met - 812 Gly**	**812**	**56.6 (820)**	**56.9 (820)**	**56.5 (820)**	**55.8 (822)**	**56.7 (822)**	**36.7 (776)**	**36.5 (775)**
2A	813 Gly - 946 Gln	134	72.6 (135)	72.6 (135)	71.9 (135)	69.6 (135)	70.4 (135)	42.6 (147)	42.7 (150)
2B	947 Ser - 1083 Gln	137	72.9 (138)	73.6 (138)	75.7 (140)	74.3 (140)	74.3 (140)	47.9 (122)	48.6 (122)
2C	1084 Gly - 1416 Glu	333	69.5 (333)	68.6 (333)	69.8 (334)	68.9 (334)	70.0 (334)	49.9 (329)	49.6 (329)
P2	**813 Pro - 1416 Glu**	**604**	**70.9 (606)**	**70.6 (606)**	**71.6 (609)**	**70.4 (609)**	**69.8 (609)**	**48.2 (598)**	**48.1 (601)**
3A	1417 Ser - 1540 Glu	124	43.9 (130)	43.1 (130)	46.2 (130)	46.2 (128)	44.6 (128)	23.2 (117)	24.0 (117)
3B	1541 Arg - 1569 Glu	29	37.9 (29)	37.9 (29)	37.9 (29)	41.4 (29)	37.9 (29)	20.7 (20)	20.7 (20)
3C	1570 Ala - 1763 Glu	194	61.1 (198)	61.6 (198)	61.1 (198)	62.8 (196)	62.2 (196)	38.9 (200)	38.9 (200)
3D	1764 Gly - 2233 Glu	470	67.5 (470)	67.7 (470)	67.7 (470)	66.0 (470)	65.3 (470)	47.9 (469)	48.3 (469)
P3	**1417 Ser - 2233 Glu**	**817**	**61.2 (827)**	**61.3 (827)**	**61.7 (827)**	**61.0 (823)**	**60.4 (823)**	**42.4 (806)**	**42.0 (806)**
Polyprotein	**1 Met - 2233 Glu**	**2233**	**56.6 (2253)**	**56.9 (2253)**	**56.5 (2256)**	**55.7 (2254)**	**56.5 (2254)**	**39.8 (2180)**	**40.2 (2182)**


(a)
1F
2F
5F
3R
4R
6R
7R
8F
9F
10R
11F
12F
13R
14R
15F
17F
16R
18R
19R
20R
(b)
Ctg 1
Ctg 2
Ctg 3
Coverage
400
300
200
100
0
(c)
5'UTR
P1
P2
P3
3'UTR
3B
Nucleotide position
755
1532
2261
3137
3191
3593
4004
5000
5375
5462
6044
7457
VP0
VP3
VP1
2A
2B
2C
3A
3C
3D
AAAn
Amino acid position
1
260
503
795
813
947
1084
1417
1541
1570
1764
2233
Q/A
Q/G
Q/G
G/P
Q/S
Q/G
E/S
E/R
E/A
E/G
Rhv-like domains
H827box/N895C
RRQ794GCR
DxExNPG812P
D1271DAGQ
S1223xxGxGKx
H1610
Y1543
D1648
G1721xC1723G
G1739xH
K1930DELR
G2057GxPSG
F2144LKR
S2061Gx3Tx3N
Y2096GDD

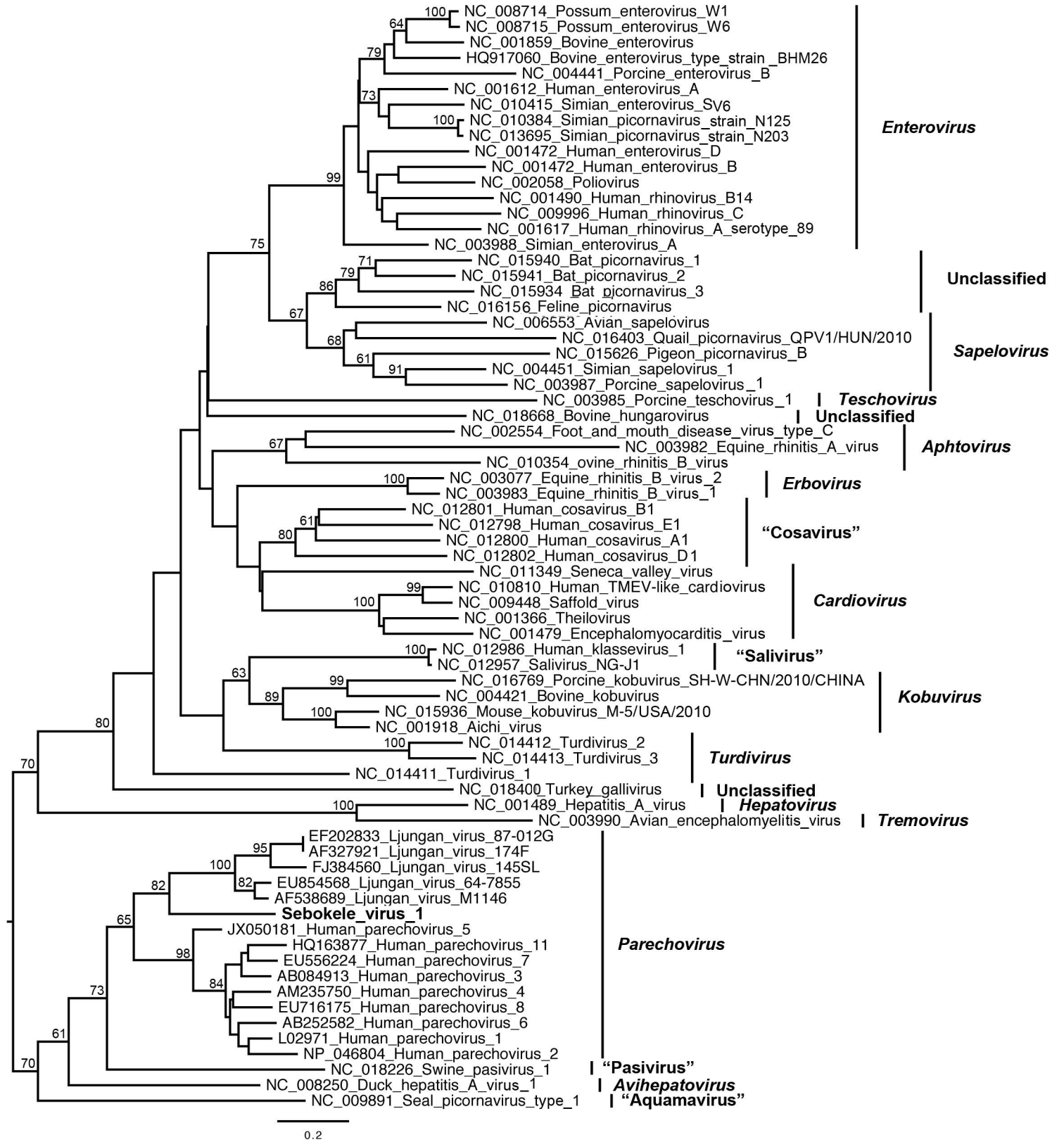

NC_008714_Possum_enterovirus_W1
NC_008715_Possum_enterovirus_W6
NC_001859_Bovine_enterovirus
HQ917060_Bovine_enterovirus_type_strain _BHM26
NC_004441_Porcine_enterovirus_B
NC_001612_Human_enterovirus_A
NC_010415_Simian_enterovirus_SV6
NC_010384_Simian_picornavirus_strain_N125
NC_013695_Simian_picornavirus_strain_N203
NC_001472_Human_enterovirus_D
NC_001472_Human_enterovirus_B
NC_002058_Poliovirus
NC_001490_Human_rhinovirus_B14
NC_009996_Human_rhinovirus_C
NC_001617_Human_rhinovirus_A_serotype_89
NC_003988_Simian_enterovirus_A
Enterovirus
NC_015940_Bat_picornavirus_1
NC_015941_Bat_picornavirus_2
NC_015934_Bat_picornavirus_3
NC_016156_Feline_picornavirus
Unclassified
NC_006553_Avian_sapelovirus
NC_016403_Quail_picornavirus_QPV1/HUN/2010
NC_015626_Pigeon_picornavirus_B
NC_004451_Simian_sapelovirus_1
NC_003987_Porcine_sapelovirus_1
Sapelovirus
NC_003985_Porcine_teschovirus_1
Teschovirus
NC_018668_Bovine_hungarovirus
Unclassified
NC_002554_Foot_and_mouth_disease_virus_type_C
NC_003982_Equine_rhinitis_A_virus
NC_010354_ovine_rhinitis_B_virus
Aphtovirus
NC_003077_Equine_rhinitis_B_virus_2
NC_003983_Equine_rhinitis_B_virus_1
Erbovirus
NC_012801_Human_cosavirus_B1
NC_012798_Human_cosavirus_E1
NC_012800_Human_cosavirus_A1
NC_012802_Human_cosavirus_D1
"Cosavirus"
NC_011349_Seneca_valley_virus
NC_010810_Human_TMEV-like_cardiovirus
NC_009448_Saffold_virus
NC_001366_Theilovirus
NC_001479_Encephalomyocarditis_virus
Cardiovirus
NC_012986_Human_klassevirus_1
NC_012957_Salivirus_NG-J1
"Salivirus"
NC_016769_Porcine_kobuvirus_SH-W-CHN/2010/CHINA
NC_004421_Bovine_kobuvirus
NC_015936_Mouse_kobuvirus_M-5/USA/2010
NC_001918_Aichi_virus
Kobuvirus
NC_014412_Turdivirus_2
NC_014413_Turdivirus_3
NC_014411_Turdivirus_1
Turdivirus
NC_018400_Turkey_gallivirus
Unclassified
NC_001489_Hepatitis_A_virus
Hepatovirus
NC_003990_Avian_encephalomyelitis_virus
Tremovirus
EF202833_Ljungan_virus_87-012G
AF327921_Ljungan_virus_174F
FJ384560_Ljungan_virus_145SL
EU854568_Ljungan_virus_64-7855
AF538689_Ljungan_virus_M1146
Sebokele_virus_1
JX050181_Human_parechovirus_5
HQ163877_Human_parechovirus_11
EU556224_Human_parechovirus_7
AB084913_Human_parechovirus_3
AM235750_Human_parechovirus_4
EU716175_Human_parechovirus_8
AB252582_Human_parechovirus_6
L02971_Human_parechovirus_1
NP_046804_Human_parechovirus_2
Parechovirus
NC_018226_Swine_pasivirus_1
"Pasivirus"
NC_008250_Duck_hepatitis_A_virus_1
Avihepatovirus
NC_009891_Seal_picornavirus_type_1
"Aquamavirus"
0.2

Table S1. List of primers used for RT-PCR amplification and Sanger sequencing of SEBV1.

Code*	Primer	Sequence (5'-3')#	Genome location¶
1	EMCV2-180-F	GGMCGAARCCGCTTGGAATA	172-192
2	SEB-805-F	GAAGAGCAGGAAACAAATGC	842-862
3	SEB-900-R	TAATGGATGCTGAAACACG	866-884
4	SEB-1050-R	ACTACCATGAGACAGCCAG	1203-1221
5	SEB-1404-F	CGTTTGAGGATTGACGTGAT	2126-2145
6	SEB-2043-R	CAACTGCTCGGAAAGTGTC	2410-2428
7	SEB-2500-R	AGCCACGATCCTTGTAGAC	3206-3224
8	SEB-3260-F	GACTCCCAGGATATACTTTC	3272-3291
9	SEB-4230-F	TAGAGTGCCAACCAATACC	4363-4381
10	SEB-4340-R	TGGTCAACATGTGGGTCAAA	4447-4466
11	SEB-5330-F	GGGTGAAGAAACATATGGTG	5223-5242
12	SEB-5480-F	TGCAGGATCTAGAGCATTGC	5472-5491
13	SEB-5500-R	CAATGCTCTAGATCCTGCAT	5471-5490
14	SEB-5810-R	TCCATAGCTGAGCCATCCA	5679-5697
15	SEB-6660-F	AGATGAGCTGCGCAAACGAG	6544-6563
16	SEB-6860-R	CATGGCAGTATGCAAGCACT	6820-6839
17	SEB-7040-F	CATCAGTTGAGTGCTTACCA	7014-7033
18	SEB-7570-R	CATAGAACAACATCCGCCTC	7426-7445
19	SEB-7640-R	CTAAAAATCAAAATCAAGGCAT	7497-7518
20	UC10	TTTTTTTTTTTTTTTTTTTTTTC	Poly(A) tail

* Code identical to that used in Fig. 1(a).
# F : Forward, R : Reverse.
¶ Numbers refer to locations in the SEBV1 genome.