



HAL
open science

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

► To cite this version:

Marie-Line Joffret, Christiane Bouchier, Marc Grandadam, Hervé Zeller, Corinne Maufrais, et al..
Genomic characterization of Sebokele virus 1 (SEBV1) reveals a new candidate species among the
genus Parechovirus.. *Journal of General Virology*, 2013, 94 (Pt_7), pp.1547-53. 10.1099/vir.0.053157-
0 . pasteur-01435337

HAL Id: pasteur-01435337

<https://pasteur.hal.science/pasteur-01435337>

Submitted on 13 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Title**

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

4

5 **Running title**

6 Molecular characterization of Sebokele virus 1

7

8 **Contents category** : Short communication

9

10 **Authors and affiliations**

11 Joffret ML^{1,2}, Bouchier C³, Grandadam M⁴, Zeller H⁵, Maufrais C⁶, Bourhy H⁷, Despres P⁸,
12 Delpeyroux F^{1,2}, Dacheux L^{7*}.

13 1 Institut Pasteur, Unité de Biologie des virus entériques, Paris, France

14 2 INSERM, U994, Paris, France

15 3 Institut Pasteur, Plate-forme Génomique, Paris, France

16 4 Institut Pasteur, Centre National de Reference des arbovirus, Unité des Interactions Moléculaires
17 Flavivirus-Hôtes, Paris, France

18 5 Institut Pasteur, Unité de Biologie des Infections Emergentes, Lyon, France

19 6 Institut Pasteur, Centre d'Informatique pour les Biologistes, Paris, France

20 7 Institut Pasteur, Unité Dynamique des Lyssavirus et Adaptation à l'Hôte, Paris, France.

21 8 Institut Pasteur, Unité des Interactions Moléculaires Flavivirus-Hôtes, Paris, France

22

23 *** Corresponding author**

24 Email : laurent.dacheux@pasteur.fr

25 Phone : +33 1 40 61 33 03

26 Fax : +33 1 40 61 30 20

27

28 **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

38 **Summary**

39 We determined the genomic features and the taxonomic classification of Sebokele virus 1 (SEBV1), a
40 previously unclassified arbovirus isolated in 1972 from rodents collected in Botambi, Central African
41 Republic. The complete genome sequence was obtained using a deep sequencing approach (Illumina
42 technology) and dedicated bioinformatics workflows for data analysis. Molecular analysis identified
43 SEBV1 as a picornavirus, most closely related to Ljungan viruses of the genus *Parechovirus*. The
44 genome has a typical Ljungan virus-like organization, including the presence of two unrelated 2A
45 protein motifs. Phylogenetic analysis confirmed that SEBV1 belongs to the parechovirus phylogroup
46 and was most closely related to the *Ljungan virus* species. However, it appeared clearly distinct from
47 all members of this phylogroup, suggesting that it represents a novel species of the genus
48 *Parechovirus*.
49

50 **Main text**

51 The family *Picornaviridae* encompasses small nonenveloped and positive single-stranded RNA
52 viruses, and includes many human and animal pathogens. The typical genome organization of
53 picornaviruses includes a single open reading frame (ORF) encoding a large protein precursor
54 (polyprotein), preceded by a 5' untranslated region (5'UTR) and followed by a 3'UTR and a
55 polyadenylated tract of variable length (Knowles *et al.*, 2012; Racaniello, 2007). The polyprotein
56 contains, in the following order, a non structural leader (L) protein in some cases, then a first domain
57 P1 encoding three structural proteins (VP0, which is autocatalytically cleaved further into VP4 and
58 VP2 in most picornaviruses, VP3 and VP1), and two distinct non structural coding regions (P2
59 encoding proteins 2A, 2B and 2C^{ATPase}, a multifunctional ATPase, and P3 encoding proteins 3A,
60 3B^{VPg}, 3C^{pro}, a cysteine protease, and 3D^{pol}, a RNA-dependent RNA polymerase). The family
61 *Picornaviridae* is currently divided into 12 genera: *Aphthovirus*, *Avihepatovirus*, *Cardiovirus*,
62 *Enterovirus*, *Erbovirus*, *Hepatovirus*, *Kobuvirus*, *Parechovirus*, *Sapelovirus*, *Senecavirus*, *Teschovirus*
63 and *Tremovirus*. However, the diversity of this viral family is far from being fully documented, and
64 numerous novel picornavirus species and genera have recently been described (Knowles *et al.*, 2012)
65 (<http://www.picornaviridae.com/>).

66 The genus *Parechovirus* includes two species, *human parechovirus* (HPeV) and *Ljungan virus* (LV),
67 each including various types and/or genotypes (Johansson *et al.*, 2002; Joki-Korpela & Hyypia, 2001;
68 Stanway & Hyypia, 1999). HPeV is a pathogen infecting mainly the gastrointestinal and the
69 respiratory tracts of children (Baumgarte *et al.*, 2008; Harvala & Simmonds, 2009; Harvala *et al.*,
70 2008; Joki-Korpela & Hyypia, 2001). Sixteen different types of HPeV and several different genotypes
71 have been identified (Knowles *et al.*, 2012) (<http://www.picornaviridae.com/>). The genome
72 organization of HPeV is typical of the family, except that the VP0 protein appears not to be cleaved
73 and its N-terminus not myristoylated (Knowles *et al.*, 2012; Stanway & Hyypia, 1999).

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

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

85 Sebokele virus 1 (SEBV1) was originally isolated in 1972 from the crushed organs of *Hylomyscus* sp.
86 (African wood mice) collected in Botambi, Central African Republic (Digoutte, 1985; El Mekki *et al.*,
87 1981; Zeller *et al.*, 1989). The virus was isolated by intracerebral inoculation of suckling mice
88 (Digoutte, 1978). The pathogenic agent appeared to be resistant to chloroform and could be filtered
89 through 220 nm-pore size filters. Attempts to identify the virus using sera against arboviruses and
90 herpesvirus, and by electron microscopy after inoculation of cell cultures, were unsuccessful. The
91 isolate was therefore described to be a non-identified arbovirus. Three similar viral isolates were
92 reported from *Muridae* in the Central African Republic: two from *Hylomyscus* sp. and one from
93 *Praomys* sp. (Digoutte, 1985).

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

100 A total of 8,509,420 reads were obtained, trimmed and mapped to the *Mus musculus* genome by using
101 CLC Assembly Cell (v. 3.11) with default parameters. Non-mapped reads (147,826 reads) were
102 extracted and assembled into contigs by three different assemblers (CLC novo assembler,
103 VelvetOptimiser and SOAP de novo). The contigs were used independently for BLASTn and
104 BLASTx identity searches against EMBL and UniProt databases, respectively. The blast2taxoclass
105 program (<http://mobyale.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://mobyale.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

134 genome sequence was complete. More than 30% of the 5'UTR sequence (nucleotide position 400 to
135 735) presented nt identities with 5'UTR sequences of parechoviruses, according to BLASTn analysis,
136 with maximum coverage (43%) and nt identity (72%) obtained with LV strain 64-7855. The 3'UTR
137 sequence of SEBV1 seems to be unique, and BLASTn searches found no related sequence. The G+C
138 content of the SEBV1 genome is 45.7%, higher than the G+C contents of LV (41.6-42.5%) and human
139 parechoviruses (39.0-39.5%). BLASTp analysis of the complete deduced polyprotein identified the
140 best matches as LV species (the best being LV strain 87-012), with 62-63% aa identity and 77-78% aa
141 similarity, and human parechoviruses HPeV-1 and -6 exhibited nearly 46% aa identity and 63% aa
142 similarity. The polyprotein sequence of SEBV1 was then aligned with its closest relatives: the Ljungan
143 viruses (strains 87-012, 174F, 145SLG, M1146 and 64-7855) and the human parechoviruses HPeV-1
144 (strain Harris) and HPeV-6 (strain NII561-2000). With NetPicoRNA prediction (Blom *et al.*, 1996),
145 this allowed the identification of potential cleavage sites and processing of this polyprotein (see Table
146 1 and Fig. 1c). The organisation of the SEBV1 genome is typical of picornaviruses, with a P1 region
147 encoding the structural proteins and two non structural regions, P2 and P3.

148 The P1 region is the genomic region of SEBV1 most divergent from the most closely related
149 picornaviruses, with nt identities of 56-57% with LV isolates and 36-37% with the most closely
150 related HPeV (Table 1). This region encodes the classical capsid proteins with, in order, VP0, VP3
151 and VP1. The P1 region contains the picornavirus capsid protein domain-like (pfam sequence cluster
152 cd00205 - Rhv-like) sequence found in other species of the genera *Parechovirus*, *Avihepatovirus* and
153 *Aquamavirus*, and other unclassified picornaviruses including turdivirus. VP0 of SEBV1 is probably
154 not cleaved into VP4 and VP2. Similarly to the other parechoviruses, the N-terminal end of this capsid
155 protein does not contain the myristylation motif (GxxxS/T) found in most picornaviruses (Chow *et al.*,
156 1987). Like LV and pasivirus 1 (SPaV1), VP0 of SEBV1 has a short N-terminal extremity (Sauvage *et al.*,
157 2012). The VP3 of SEBV1 has a long N-terminal extremity enriched with basic residues, like
158 those of parechoviruses (Johansson *et al.*, 2002). However, the highly conserved KxKxxRxK motif in
159 the parechovirus P1 region was not found, and is replaced by the motif R₂₇₀FKWTRNN (Johansson *et al.*,
160 2002; Williams *et al.*, 2009). VP1 is predicted to be cleaved at the canonical cleavage site
161 DxExNPG₈₁₂P (Fig. 1c) (Ryan & Flint, 1997). The parechovirus (PS)ALxAxETG motif was not

162 present in this capsid protein. As in LV, SPaV1 and some HPeVs, SEBV1 lacked the RGD tri-peptide
163 involved in integrin receptor binding, raising questions about the mechanism of cellular entry and the
164 host tropism of this virus (Knowles *et al.*, 2012; Merilahti *et al.*, 2012; Sauvage *et al.*, 2012; Williams
165 *et al.*, 2009). The VP1 protein of SEBV1, like those of LV and SPaV1, contained one N-terminal
166 insertion (8 aa long) and a unique C terminal extension (48 aa long). Similarly to LV, this protein
167 contains the 3C^{pro} cleavage site RRQ₇₉₄GCR predicted by NetPicoRNA at the end of its N-terminus,
168 suggesting the existence of a short 2A-like protein (18 aa) (Fig. 1c).

169 The non structural P2 region of SEBV1 shows 70-72% aa sequence identity with LV and nearly 48%
170 with HPeV. This region was deduced to encompass the non-structural proteins 2A, 2B and 2C (Table
171 1 and Fig 1c). The 2A protein possesses the highly conserved picornaviral H-box/NC regions,
172 involved in the control of cell proliferation (Hughes & Stanway, 2000). The absence of the conserved
173 GxCG motif suggests that the 2A protein of SEBV1 does not have trypsin-like proteolytic activity
174 (Lamphear *et al.*, 1993). The 2C protein, like those of other parechoviruses, carries the SxxGxGKx
175 NTase motif and the D₁₂₇₁DLxQ helicase motif (Gorbalenya *et al.*, 1989a).

176 The P3 region encompasses the non-structural proteins 3A, 3B, 3C and 3D, displaying 60-62% amino
177 acid identity with LV and nearly 42% with HPeV (Table 1 and Fig 1c). The protein 3B contains the
178 conserved tyrosine (Y) residue at the third position of the predicted N-terminus, present in all known
179 picornaviruses and necessary for the priming function of the VPg (Ambros & Baltimore, 1978). The
180 conserved catalytic triad H₁₆₁₀-D₁₆₄₈-C₁₇₂₃ identifies the 3C protein as a cysteine-protease (Bazan &
181 Fletterick, 1988) with the active site motif G₁₇₂₁xCG and G₁₇₃₉xH (Gorbalenya *et al.*, 1989b). The 3D
182 protein contained the five highly conserved RNA polymerase motifs: K₁₉₃₀DELRL, G₂₀₅₇GxPSG,
183 S₂₀₆₁GX₃TX₃N, Y₂₀₉₆GDD and F₂₁₄₄LKR (Koonin, 1991).

184

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

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

192

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

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

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

213

214 **Legends to the figures**

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

224

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

229

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

234

235 **Acknowledgements**

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

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

240 **References**

- 241 **Ambros, V. & Baltimore, D. (1978).** Protein is linked to the 5' end of poliovirus RNA by
242 a phosphodiester linkage to tyrosine. *The Journal of biological chemistry* **253**,
243 5263-5266.
- 244 **Baumgarte, S., de Souza Luna, L. K., Grywna, K., Panning, M., Drexler, J. F., Karsten,**
245 **C., Huppertz, H. I. & Drosten, C. (2008).** Prevalence, types, and RNA
246 concentrations of human parechoviruses, including a sixth parechovirus type, in
247 stool samples from patients with acute enteritis. *Journal of clinical microbiology*
248 **46**, 242-248.
- 249 **Bazan, J. F. & Fletterick, R. J. (1988).** Viral cysteine proteases are homologous to the
250 trypsin-like family of serine proteases: structural and functional implications.
251 *Proceedings of the National Academy of Sciences of the United States of America*
252 **85**, 7872-7876.
- 253 **Blom, N., Hansen, J., Blaas, D. & Brunak, S. (1996).** Cleavage site analysis in
254 picornaviral polyproteins: discovering cellular targets by neural networks.
255 *Protein science : a publication of the Protein Society* **5**, 2203-2216.
- 256 **Chow, M., Newman, J. F., Filman, D., Hogle, J. M., Rowlands, D. J. & Brown, F. (1987).**
257 Myristylation of picornavirus capsid protein VP4 and its structural significance.
258 *Nature* **327**, 482-486.
- 259 **Dacheux, L., Berthet, N., Dissard, G., Holmes, E. C., Delmas, O., Larrous, F., Guigon,**
260 **G., Dickinson, P., Faye, O. & other authors (2010).** Application of broad-
261 spectrum resequencing microarray for genotyping rhabdoviruses. *Journal of*
262 *virology* **84**, 9557-9574.
- 263 **Digoutte, J. P. (1978).** Sebokele (SEB). *The American journal of tropical medicine and*
264 *hygiene* **17**, 424-425.
- 265 **Digoutte, J. P., Germain, M (1985).** Sebokele (An B 1227d), nouvel Arbovirus non
266 groupé. In *International Catalogue of Arboviruses Including Certain Other Virus of*
267 *Vertebrates*, Third edn, pp. 917-918. Edited by N. Karabatsos. San Antonio, Texas:
268 Am. Soc. Trop. Med. Hyg.
- 269 **Ekstrom, J. O., Tolf, C., Fahlgren, C., Johansson, E. S., Arbrandt, G., Niklasson, B.,**
270 **Edman, K. A. & Lindberg, A. M. (2007).** Replication of Ljungan virus in cell
271 culture: the genomic 5'-end, infectious cDNA clones and host cell response to
272 viral infections. *Virus research* **130**, 129-139.
- 273 **El Mekki, A. A., Nieuwenhuysen, P., van der Groen, G. & Pattyn, S. R. (1981).**
274 Characterization of some ungrouped viruses. *Transactions of the Royal Society of*
275 *Tropical Medicine and Hygiene* **75**, 799-806.
- 276 **Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989a).** Two
277 related superfamilies of putative helicases involved in replication, recombination,
278 repair and expression of DNA and RNA genomes. *Nucleic acids research* **17**, 4713-
279 4730.
- 280 **Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989b).** Cysteine
281 proteases of positive strand RNA viruses and chymotrypsin-like serine proteases.
282 A distinct protein superfamily with a common structural fold. *FEBS letters* **243**,
283 103-114.
- 284 **Hall, T. (1999).** BioEdit: a user-friendly biological sequence alignment editor and
285 analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95-98.
- 286 **Harvala, H. & Simmonds, P. (2009).** Human parechoviruses: biology, epidemiology
287 and clinical significance. *Journal of clinical virology : the official publication of the*
288 *Pan American Society for Clinical Virology* **45**, 1-9.

- 289 **Harvala, H., Robertson, I., McWilliam Leitch, E. C., Benschop, K., Wolthers, K. C.,**
290 **Templeton, K. & Simmonds, P. (2008).** Epidemiology and clinical associations
291 of human parechovirus respiratory infections. *Journal of clinical microbiology* **46**,
292 3446-3453.
- 293 **Hughes, P. J. & Stanway, G. (2000).** The 2A proteins of three diverse picornaviruses are
294 related to each other and to the H-rev107 family of proteins involved in the
295 control of cell proliferation. *The Journal of general virology* **81**, 201-207.
- 296 **Johansson, E. S., Niklasson, B., Tesh, R. B., Shafren, D. R., Travassos da Rosa, A. P. &**
297 **Lindberg, A. M. (2003).** Molecular characterization of M1146, an American
298 isolate of Ljungan virus (LV) reveals the presence of a new LV genotype. *The*
299 *Journal of general virology* **84**, 837-844.
- 300 **Johansson, E. S., Ekstrom, J. O., Shafren, D. R., Frisk, G., Hyypia, T., Edman, K. &**
301 **Lindberg, A. M. (2004).** Cell culture propagation and biochemical analysis of the
302 Ljungan virus prototype strain. *Biochemical and biophysical research*
303 *communications* **317**, 1023-1029.
- 304 **Johansson, S., Niklasson, B., Maizel, J., Gorbalenya, A. E. & Lindberg, A. M. (2002).**
305 Molecular analysis of three Ljungan virus isolates reveals a new, close-to-root
306 lineage of the Picornaviridae with a cluster of two unrelated 2A proteins. *Journal*
307 *of virology* **76**, 8920-8930.
- 308 **Joki-Korpela, P. & Hyypia, T. (2001).** Parechoviruses, a novel group of human
309 picornaviruses. *Annals of medicine* **33**, 466-471.
- 310 **Knowles, N. J., Hovi, T., Hyypia, T., King, A. M. Q., Lindberg, A. M., Pallansch, M. A.,**
311 **Palmenberg, A. C., Simmonds, P., Skern, T. & other authors (2012).**
312 Picornaviridae. In *Virus taxonomy: classification and Nomenclature of Viruses:*
313 *Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 855-880.
314 Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. San Diego:
315 Elsevier Academic Press.
- 316 **Koonin, E. V. (1991).** The phylogeny of RNA-dependent RNA polymerases of positive-
317 strand RNA viruses. *The Journal of general virology* **72 (Pt 9)**, 2197-2206.
- 318 **Lamphear, B. J., Yan, R., Yang, F., Waters, D., Liebig, H. D., Klump, H., Kuechler, E.,**
319 **Skern, T. & Rhoads, R. E. (1993).** Mapping the cleavage site in protein synthesis
320 initiation factor eIF-4 gamma of the 2A proteases from human Coxsackievirus
321 and rhinovirus. *The Journal of biological chemistry* **268**, 19200-19203.
- 322 **Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A.,**
323 **McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007).**
324 Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
- 325 **Merilahti, P., Koskinen, S., Heikkila, O., Karelehto, E. & Susi, P. (2012).** Endocytosis
326 of integrin-binding human picornaviruses. *Advances in virology* **2012**, 547530.
- 327 **Niklasson, B., Hornfeldt, B. & Lundman, B. (1998).** Could myocarditis, insulin-
328 dependent diabetes mellitus, and Guillain-Barre syndrome be caused by one or
329 more infectious agents carried by rodents? *Emerging infectious diseases* **4**, 187-
330 193.
- 331 **Niklasson, B., Samsioe, A., Papadogiannakis, N., Kawecky, A., Hornfeldt, B., Saade, G.**
332 **R. & Klitz, W. (2007).** Association of zoonotic Ljungan virus with intrauterine
333 fetal deaths. *Birth defects research Part A, Clinical and molecular teratology* **79**,
334 488-493.
- 335 **Niklasson, B., Kinnunen, L., Hornfeldt, B., Horling, J., Benemar, C., Hedlund, K. O.,**
336 **Matskova, L., Hyypia, T. & Winberg, G. (1999).** A new picornavirus isolated
337 from bank voles (*Clethrionomys glareolus*). *Virology* **255**, 86-93.

- 338 **Racaniello, V. (2007).** Picornaviridae: the viruses and their replication. In *Fields*
339 *Virology*, Fifth edn, pp. 795-838. Edited by D. M. Knipe & P. M. Howley.
340 Philadelphia, PA: Lippincott Williams and Wilkins.
- 341 **Ryan, M. D. & Flint, M. (1997).** Virus-encoded proteinases of the picornavirus super-
342 group. *The Journal of general virology* **78 (Pt 4)**, 699-723.
- 343 **Sauvage, V., Ar Gouilh, M., Cheval, J., Muth, E., Pariente, K., Burguiere, A., Caro, V.,**
344 **Manuguerra, J. C. & Eloit, M. (2012).** A member of a new Picornaviridae genus
345 is shed in pig feces. *Journal of virology* **86**, 10036-10046.
- 346 **Stanway, G. & Hyypia, T. (1999).** Parechoviruses. *Journal of virology* **73**, 5249-5254.
- 347 **Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007).** MEGA4: Molecular Evolutionary
348 Genetics Analysis (MEGA) software version 4.0. *Molecular biology and evolution*
349 **24**, 1596-1599.
- 350 **Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997).**
351 The CLUSTAL_X windows interface: flexible strategies for multiple sequence
352 alignment aided by quality analysis tools. *Nucleic acids research* **25**, 4876-4882.
- 353 **Tolf, C., Gullberg, M., Johansson, E. S., Tesh, R. B., Andersson, B. & Lindberg, A. M.**
354 **(2009).** Molecular characterization of a novel Ljungan virus (Parechovirus;
355 Picornaviridae) reveals a fourth genotype and indicates ancestral recombination.
356 *The Journal of general virology* **90**, 843-853.
- 357 **Tolf, C., Ekstrom, J. O., Gullberg, M., Arbrandt, G., Niklasson, B., Frisk, G., Liljeqvist,**
358 **J. A., Edman, K. & Lindberg, A. M. (2008).** Characterization of polyclonal
359 antibodies against the capsid proteins of Ljungan virus. *Journal of virological*
360 *methods* **150**, 34-40.
- 361 **Whitney, E., Roz, A. P. & Rayner, G. A. (1970).** Two viruses isolated from rodents
362 (*Clethrionomys gapperi* and *Microtus pennsylvanicus*) trapped in St. Lawrence
363 County, New York. *Journal of wildlife diseases* **6**, 48-55.
- 364 **Williams, C. H., Panayiotou, M., Girling, G. D., Peard, C. I., Oikarinen, S., Hyoty, H. &**
365 **Stanway, G. (2009).** Evolution and conservation in human parechovirus
366 genomes. *The Journal of general virology* **90**, 1702-1712.
- 367 **Zeller, H. G., Karabatsos, N., Calisher, C. H., Digoutte, J. P., Murphy, F. A. & Shope, R.**
368 **E. (1989).** Electron microscopy and antigenic studies of uncharacterized viruses.
369 I. Evidence suggesting the placement of viruses in families Arenaviridae,
370 Paramyxoviridae, or Poxviridae. *Archives of virology* **108**, 191-209.

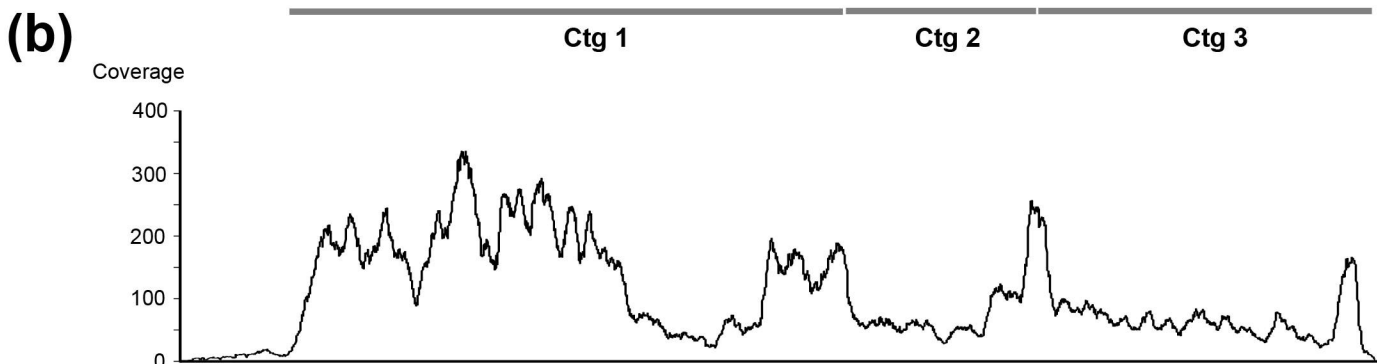
373

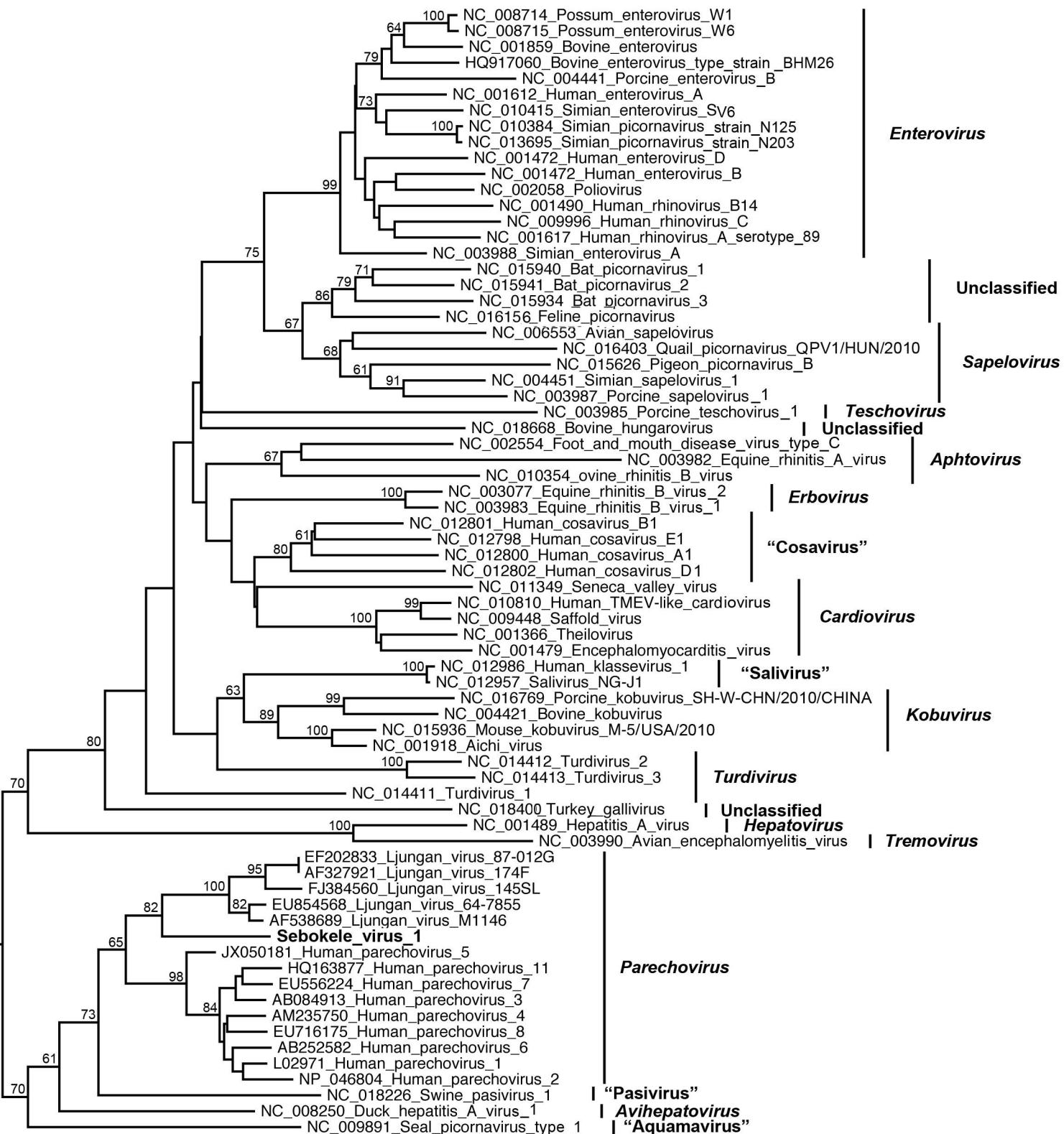
374
375
376

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)

377
378





0.2

1
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	GGMCGAARCCGCTTGAATA	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

3
4
5
6
7

* Code identical to that used in Fig. 1(a).

F : Forward, R : Reverse.

¶ Numbers refer to locations in the SEBV1 genome.