Metagenomic Analysis of Shrew Enteric Virome Reveals Novel Viruses Related to Human Stool-Associated Viruses

Michihito Sasaki[1], Yasuko Orba[1], Keisuke Ueno[2], Akihiro Ishii[3], Ladslav Moonga[4], Bernard M. Hang'ombe[4], Aaron S. Mweene[5], Kimihito Ito[2] and Hirofumi Sawa[1,6*]

[1]Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, N20, W10, Kita-ku, Sapporo 001-0020, Japan

[2]Division of Bioinformatics, Research Center for Zoonosis Control, Hokkaido University, N20, W10, Kita-ku, Sapporo 001-0020, Japan

[3]Hokudai Center for Zoonosis Control in Zambia, PO Box 32379, Lusaka, Zambia

[4]Department of Paraclinical Studies, School of Veterinary and Medicine, University of Zambia, PO Box 32379, Lusaka, Zambia

[5]Department of Disease Control, School of Veterinary and Medicine, University of Zambia, PO Box 32379, Lusaka, Zambia

[6]Global Institution for Collaborative Research and Education, Hokkaido University, N20, W10, Kita-ku, Sapporo 001-0020, Japan

[*]Corresponding author:
Hirofumi Sawa
Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, N20, W10, Kita-ku, Sapporo 001-0020, Japan
Tel: +81-11-706-5185, Fax: +81-11-706-7370, E-mail: h-sawa@czc.hokudai.ac.jp

Running title: Shrew Enteric Virome

Contents Category: Other viruses

Word count of the summary: 181 words

Word count of the main text: 5360 words

Number of tables and figures: 3 tables and 4 figures

Number of supplementary figures: 1 supplementary figure

Footnote: The GenBank/EMBL/DDBJ accession numbers for the viral sequences reported in this paper are AB937980-AB937987 and AB937989-AB937991. DRA002561 is the GenBank/EMBL/DDBJ accession number for the raw sequence reads from the metagenomic library.

Summary

Shrews are small insectivorous mammals that are distributed worldwide. Similar to rodents, shrews live on the ground and are commonly found near human residences. In this study, we investigated the enteric virome of wild shrews in the genus *Crocidurinae* using a sequence-independent viral metagenomics approach. A large portion of the shrew enteric virome was composed of insect viruses, while novel viruses including cyclovirus, picornavirus and picorna-like virus were also identified. Several cycloviruses, including variants of human cycloviruses detected in cerebrospinal fluid (CSF) and stool, were detected in wild shrews at a high prevalence rate. The identified picornavirus is distantly related to human parechovirus, inferring the presence of a new genus in this family. The identified picorna-like viruses were characterized as different species of calhevirus 1, which was previously discovered in human stool. Complete or nearly complete genome sequences of these novel viruses were determined in this study and then were subjected to further genetic characterization. Our study provides an initial view of the diversity and distinctiveness of the shrew enteric virome and highlights unique novel viruses related to human stool-associated viruses.

51 **Introduction**

52 Shrews are small, mole-like insectivorous mammals in the family *Soricidae*, order
53 *Soricomorpha*. Members of this family comprise at least 385 species with a nearly global
54 distribution (Wilson & Reeder, 2011). In Africa, *Crocidura* spp. are distributed throughout
55 the continent, including in Zambia (Dubey *et al.*, 2007). Despite their similarity in size and
56 appearance to rodents (the order *Rodentia*), they are genetically closer to bats (the order
57 *Chiroptera*) than to rodents (Guo *et al.*, 2013). Similar to rodents, shrews live on the ground
58 and are commonly found near human residences. Although rodents and bats are
59 well-known reservoirs of a number of zoonotic infectious diseases (Meerburg *et al.*, 2009;
60 Smith & Wang, 2013), the knowledge of pathogens harbored by shrews is comparatively
61 limited. Shrews are reservoirs of borna disease virus and a number of hantavirus species
62 (Dürrwald *et al.*, 2014; Hilbe *et al.*, 2006; Witkowski *et al.*, 2014; Yanagihara *et al.*, 2014).
63 Paramyxoviruses closely related to henipavirus have also been found in shrews (Sasaki *et*
64 *al.*, 2014). Collectively, these studies suggest the unique viral diversity of shrews carries
65 potential risk for public health.

66 Mammals are speculated to harbor at least 320,000 undiscovered viruses (Anthony *et*
67 *al.*, 2013). Most emerging diseases in humans are caused by unexpected transmission by
68 the microbial flora of wildlife and domestic animals. Therefore, to predict and manage
69 future outbreaks, it is helpful to investigate the baseline level of viruses in animals and
70 virus-host relationships (Mokili *et al.*, 2012; Morse *et al.*, 2012). The advent of
71 high-throughput sequencing technology has enabled comprehensive approaches for the
72 simultaneous detection of many viral genomes and the identification of unknown viral
73 genomes without viral isolation (Firth & Lipkin, 2013). Using high-throughput sequencing,
74 viral metagenomics approaches have elucidated shown enteric viromes, resulting in the

75 discovery of unknown viruses in a variety of mammals, including nonhuman primates, bats,
76 pigs, rodents, cats, sea lions, martens, badgers, foxes, ferrets and pigeons (Baker *et al.*,
77 2013; Bodewes *et al.*, 2013; Dacheux *et al.*, 2014; Donaldson *et al.*, 2010; Ge *et al.*, 2012;
78 Handley *et al.*, 2012; Li *et al.*, 2011b; Li *et al.*, 2010b; Ng *et al.*, 2014; Phan *et al.*, 2011;
79 Phan *et al.*, 2013a; Shan *et al.*, 2011; Smits *et al.*, 2013a; van den Brand *et al.*, 2012; Wu *et*
80 *al.*, 2012). Further molecular characterization has revealed high nucleotide sequence
81 diversity and unique genome organization of novel viruses (Boros *et al.*, 2013; Boros *et al.*,
82 2012; Li *et al.*, 2010a; Phan *et al.*, 2013b; Sauvage *et al.*, 2012). To the best of our
83 knowledge, no report of the shrew enteric virome has been described.
84 In the present study, we aimed to investigate the enteric viral flora of wild shrews
85 living in close proximity to human habitation. Using a viral metagenomics approach, we
86 identified novel viruses related to human stool-associated viruses. These viruses were
87 subjected to further genetic characterization.
88

89 **Results**

90 **Sequence data overview**

91 Viral nucleic acids were isolated from a combined suspension of intestinal contents
92 from 22 *Crocidura hirta* and 1 *Crocidura luna* captured at Mpulungu in the Northern
93 Province of Zambia. Before the extraction of nucleic acids, the intestinal contents
94 suspension was filtered and treated with nucleases to reduce incorporation of nucleic acids
95 derived from the host and/or bacteria. For the sequencing of RNA viruses, cDNA was
96 prepared from the isolated viral nucleic acids by reverse transcription (RT).
97 High-throughput sequencing generated a total of 6,243,181 reads with an average length of
98 266 bp and a range of lengths from 8 to 624 bp. Sequence reads were compared with the
99 NCBI nucleotide database (nt) by using BLASTN. The taxonomic content of the sequences
100 was computed by the lowest common ancestor method in MEGAN (Huson *et al.*, 2007). As
101 a result, 893,430 reads in total were assigned to taxonomic groups and 726,286 reads
102 (81.2% of all the assigned sequence reads) were assigned as virus-related sequences, while
103 124,699 reads (14.0%) and 22,999 reads (2.6%) were related to bacteria and eukaryota,
104 respectively (Fig. 1a), indicating that viral nucleic acids were enriched by the filtration and
105 nuclease treatment.

106 Among virus-related sequence reads, 50.0% of the reads were grouped into
107 single-stranded RNA viruses, most of which were assigned to the family *Dicistroviridae*
108 (Fig. 1b), a family of invertebrate viruses. In addition, 15.7% of the reads were
109 double-stranded DNA viruses, the majority of which belonged to the bacteriophage families
110 *Siphoviridae* (8.7%), *Myoviridae* (4.0%), and *Podoviridae* (2.7%). A total of 32% of the
111 reads were single-stranded DNA viruses from the family *Parvoviridae* (18.6%) and
112 *Circoviridae* (13.8%). More than 99% of the reads assigned to the *Parvoviridae* family

were densoviruses, insect and crustacean parvoviruses. Very few sequences (<0.1%) corresponding to Drosophila A virus, which is a known species of double-stranded RNA virus, were detected in this experiment. A large proportion of the obtained sequences obtained were invertebrate viruses (67.9%, Fig. 1c).

No viruses with relatively high sequence identity (>90%) to known mammalian viruses were identified in this analysis. Sequences with relatively low similarity to known mammalian viruses such as human cyclovirus (*Circoviridae*), human parechovirus (*Picornaviridae*) and calhevirus 1 (CHV1, unclassified picorna-like virus) were detected and attributed to novel mammalian viruses. These sequence reads were assembled into several contigs by *de novo* assembly, but these contigs did not cover the overall genome sequence. Therefore, we bridged these sequence reads and contigs by conventional PCR or RT-PCR and confirmed the sequences by Sanger sequencing. The full or nearly full genome sequences of these viruses were determined and further characterized.

Identification of novel cycloviruses

Cycloviruses are members of the newly proposed genus *Cyclovirus* within the family *Circoviridae* (King *et al.*, 2012; Li *et al.*, 2010a). They have a circular single-stranded DNA genome encoding a capsid protein (Cap) gene on the virion sense and a rolling circle replication initiator protein (Rep) gene on the complementary sense (Rosario *et al.*, 2012). On the basis of the cyclovirus-related sequence reads detected by our metagenomic analysis, we designed primers and amplified the complete genome of cycloviruses by inverse PCR. Notably, cyclovirus genomes were identified in the intestinal content of 91% (21/23) of the sampled shrews (Table 1), and dual detection of different cycloviruses was observed in three samples. Based on sequence identity, the cycloviruses we identified can be grouped

137 into the following five types: ZM01, ZM41, ZM36a, ZM50a and ZM62; each of these
138 consisted of isolates sharing more than 95% nucleotide identity with the representative
139 isolates.

140 The genome organization of the representative isolates is shown in Fig. 2a. All
141 identified cyclovirus genomes, which ranged from 1,851 nucleotides to 1,865 nucleotides,
142 contained ORFs encoding Cap, Rep and hypothetical proteins. The potential splice acceptor
143 sequence (TTG↓GT) and donor sequence (CAG↓CA) was observed in the Rep coding
144 region of all identified cycloviruses. Similar to known cycloviruses, the Rep proteins of the
145 identified viruses had three conserved rolling circle amplification (RCA) motifs; RCA I
146 (WTLNN), RCA II (HLQGFCNL) and RCA III (YCSKGGD). They also had three
147 conserved superfamily 3 helicase motifs; the Walker A (GCTGTGKS), B (VVIDDFYGW)
148 and C (ITSE) motifs (Dayaram *et al.*, 2013). A putative stem-loop structure containing a
149 highly conserved nonamer sequence (TAGTATTAC) is considered the origin of replication
150 and was identified at the intergenic region between the 5′ ends of the Cap and Rep ORFs
151 (Figs. 2a and 2b).

152 Phylogenetic analysis was performed on the basis of the amino acid sequences of
153 full-length Rep (Fig. 2c). All identified cycloviruses clustered phylogenetically with human
154 cyclovirus CyCV-VN and human cyclovirus VS5700009, which were initially identified in
155 cerebrospinal fluid (CSF) from patients with suspected central nervous system (CNS)
156 infections or unexplained paraplegia (Smits *et al.*, 2013b; Tan *et al.*, 2013). CyCV-VN has
157 also been detected in the stools of healthy children (Tan *et al.*, 2013). The International
158 Committee on Taxonomy of Viruses (ICTV) suggests criteria for circovirus species
159 demarcation of genome nucleotide identities of less than 75% and Cap protein amino acid
160 identities of less than 70% (King *et al.*, 2012). Accordingly, all isolates except for ZM36a

may be variants of CyCV-VN (Table 2). Although the isolate ZM36a exhibits relatively low sequence identity with known cyclovirus species, our phylogenetic study showed that ZM36a fell inside a cluster containing other cycloviruses we identified. Therefore, ZM36a would be considered the same species of cyclovirus as the other viruses in this cluster.

Identification of a novel picornavirus

Parechovirus is a genus in the family *Picornaviridae* that comprises the following 3 species: human parechovirus, Ljungan virus and Sebokele virus. Human parechovirus is a common enteric pathogen associated with gastroenteritis, respiratory illness and, rarely, more severe diseases such as myocarditis, encephalitis, pneumonia, meningitis and flaccid paralysis (Esposito *et al.*, 2014). Ljungan virus and Sebokele virus were isolated from rodents in Sweden (Niklasson *et al.*, 1999) and the Central African Republic (Joffret *et al.*, 2013), respectively. We identified sequence reads distantly related to known parechoviruses and temporarily named this *Crocidura hirta*-derived picornavirus “Crohivirus 1 (CroV1)”.

In general, picornaviruses have single-stranded RNA genomes encoding a single polyprotein downstream of an internal ribosome entry site (IRES) element. A nearly complete genome sequence of CroV1, 7,321 nucleotides in length, was determined, but the 5′-end sequence was not obtained by 5′ Rapid amplification of cDNA end (RACE) experiments (Fig. 3a). A single large ORF encoding a putative polyprotein of 2,170 amino acids and the initiation codon (AUG) was located in the Kozak consensus sequence (AAGAUGG) in the CroV1 genome. Potential polyprotein cleavage sites were predicted by the NetPicoRNA program (Blom *et al.*, 1996) and multiple alignment with members of the genus *Parechovirus* (Fig. 3a). Comparison analyses of amino acid sequences identified several distinctive motifs conserved across the different genera in the family

185 *Picornaviridae* (Le Gall *et al.*, 2008). The P1 region of CroV1 contains putative capsid
186 proteins with the characteristic motif KxKxxRxK (x = all amino acid residues), which is
187 conserved in human parechoviruses and Ljungan viruses but not Sebokele virus (Williams
188 *et al.*, 2009). The P2 and P3 regions contain non-structural proteins involved in protein
189 processing and genome replication. The ribosomal skipping 2A sequence (DxExNPGP)
190 was identified in the N-terminal P2 region of CroV1 as well as Ljungan virus and Sebokele
191 virus (Luke *et al.*, 2008). The picornavirus 2C protein belongs to the superfamily 3
192 helicases, and the 2C protein of CroV1 contains the conserved walker motifs (GxxGxGKS
193 and DD) critical for the ATPase activity of the 2C protein (Sweeney *et al.*, 2010).
194 Consistent with all other picornaviruses, the tyrosine residue (Y) was present at position 3
195 of the predicted N-terminus of the 3B protein (Vpg, viral genome-linked protein); this
196 residue is responsible for the covalent linkage of Vpg to the 5′ end of the viral RNA
197 genome (Goodfellow, 2011). The 3C protease harbors the catalytic triad H-D-C and the
198 conserved protease active sites GxCG and GxH (Gorbalenya *et al.*, 1989). The 3D
199 RNA-dependent RNA polymerase (RdRp) also harbors the highly conserved KDELR,
200 GxPSG, YGDD and FLKR motifs (Kamer & Argos, 1984). The positions of these identified
201 motifs are mapped on the diagram of the CroV1 genome organization shown in Fig. 3a.
202 SimPlot sliding window analysis revealed that 3D RdRp is relatively conserved
203 between CroV1 and parechovirus species, while a high degree of amino acid divergence
204 was observed in the P1 region (Fig. 3b). The pairwise amino acid identities of the P1, P2
205 and P3 regions of CroV1 and those of its closest relatives were as follows: 33.9% identity
206 of the P1 region with Human parechovirus type 3, 38.2% identity of the P2 region with
207 Sebokele virus 1, and 39.7% identity of the P3 region with Ljungan virus strain M1146
208 (Table 3). According to the taxonomy guidelines of the ICTV Picornaviridae Study Group

209 (http://www.picornastudygroup.com/definitions/genus_definition.htm), members of a
210 picornavirus genus share greater than 40%, 40% and 50% amino acid identity in the P1, P2
211 and P3 regions, respectively. Therefore, CroV1 is not assigned to any genus and may be a
212 member of a new picornavirus genus. Phylogenetic analysis of 3D RdRp revealed a clear
213 phylogenetic division between CroV1 and parechoviruses (Fig. 3c). CroV1 was also
214 distinct from ferret parechovirus, a recently discovered picornavirus distantly related to
215 parechoviruses (Smits *et al.*, 2013a), and clustered with Swine pasivirus 1 and PLV-CHN,
216 which are new picornaviruses identified in piglets (Sauvage *et al.*, 2012; Yu *et al.*, 2013).
217
218 **Identification of novel picorna-like viruses**
219 In the analysis of sequence reads from high-throughput sequencing, we identified two
220 similar picorna-like virus sequences related to CHV1, a recently identified unclassified
221 picorna-like virus identified in the feces of a patient with acute flaccid paralysis (Kapoor *et*
222 *al.*, 2010). We named the viruses calhevirus 2a (CHV2a, 9,837 nucleotides) and calhevirus
223 2b (CHV2b, 8,899 nucleotides). We determined large proportions of the viral genome
224 organization, including a partial ORF1 encoding a putative nonstructural polyprotein, an
225 intergenic region, an ORF2 encoding a putative structural protein, a putative ORF3 with
226 unknown function and the 3′ untranslated region (UTR) (Fig. 4a). Consistent with CHV1,
227 the putative nonstructural protein had the following characteristic motifs: a Walker A motif
228 (GxxGxGKS) and B motif (DD) for helicase activity, an H-D-S motif for protease activity
229 and highly conserved RdRp motifs (KDELR, YGDD, FLKR) (Le Gall *et al.*, 2008). Similar
230 genome organizations have been observed in dicistroviruses, which are pathogenic
231 picorna-like insect viruses (Bonning & Miller, 2010). In the dicistrovirus genome, the IRES
232 element is present in the intergenic region between two ORFs and is characterized by

multiple stem-loops and pseudoknots (Nakashima & Uchiumi, 2009). Although a relatively longer intergenic region was present in the genomes of CHV2a and CHV2b, none of the conserved motifs of the dicistrovirus IRES element were observed.

A BLASTP search revealed that only the putative RdRp regions of CHV2a and CHV2b shared low amino acid sequence identity with members of the order *Picornavirales*. Therefore, phylogenetic analysis was conducted based on the RdRp region. CHV2a and CHV2b were closely related to CHV1 but distinct from all other known picorna-like viruses (Fig. 4b).

To infer the possible host(s) for CHV2a and CHV2b, we performed nucleotide composition analysis and subsequent canonical discriminant analysis (Kapoor *et al.*, 2010; Shan *et al.*, 2011). Analysis of the mononucleotide and dinucleotide frequencies of the viral genomes suggested that CHV2a and CHV2b, as well as CHV1, originated from arthropod hosts (Fig. S1, available in the online Supplementary Material).

Molecular screening of the viruses identified in intestinal contents and tissue samples

In addition to the aforementioned cyclovirus screening of shrew intestinal contents, we performed RT-PCR screening of the same samples to identify RNA viruses. The results are summarized in Table 1. CroV1, CHV2a and CHV2b were detected in 4-17% of the intestinal contents from the individual shrews. We further evaluated the presence of each virus in the lung, liver, spleen and kidney tissues of shrews showing a positive result in the screening test on the intestinal contents. CroV1 were detected in the liver and spleen as well as the intestinal contents. None of the other viruses were detected in tissue samples.

We then applied the viral screening test to rodent samples obtained at the same sampling occasion. Only CHV2b and various types of cycloviruses were detected in the

257 intestinal contents but not in rodent tissues (Table 1). Dual detection of different
258 cycloviruses was observed in the intestinal contents of two rodents. Of the 20 cyclovirus
259 sequences obtained from the intestinal contents of 18 rodents, 18 sequences corresponded
260 to cycloviruses ZM01, ZM41, ZM36a, ZM50a and ZM62, which were identified in shrew
261 intestinal contents in this study and described above. The remaining two identical
262 sequences, named cyclovirus ZM32, shared 92% nucleotide sequence identity with ZM50a
263 (Table 2) and were included in the phylogenetic analysis of cycloviruses (Fig. 2c).
264 Commercial columns and reagents can be unexpectedly contaminated with nucleic
265 acids, including circovirus-like sequences (Lysholm *et al.*, 2012). To exclude the possibility
266 of false detection of viruses via contamination, the sample lysis buffers from each of the
267 processed nucleic acid extraction kits were used as negative control specimens. No positive
268 signal was detected from these controls in our molecular screening experiments.
269

Discussion

Insect viruses constituted a large proportion of the shrew enteric virome and mainly included members of *Dicistroviridae* and *Densovirinae*. This result reflects the diets of shrews as well as insectivorous bats (Donaldson *et al.*, 2010; Ge *et al.*, 2012; Li *et al.*, 2010b). Although it remains unclear whether the novel viruses described in this study infect shrews, the detection of CroV1 from some tissues supports the hypothesis of replication in the organs of shrews (Delwart, 2013).

Cycloviruses ZM01, ZM41, ZM36a, ZM50a and ZM62 were detected in the intestinal contents of both shrews and rodents, suggesting circulation of these cycloviruses between the shrew and rodent populations. By contrast, CroV1 and CHV2a were identified in the intestinal contents of shrews but not rodents. Nevertheless, considering the influence of some biases such as the small size of the population and limited geography, the host ranges of the identified viruses remain to be determined. In this study, we also cannot exclude the possibility that the cycloviruses we identified came from common prey. Further epidemiological studies are necessary to understand the distribution and host specificity of these viruses.

Recent metagenomic studies have identified a number of cycloviruses from the feces, respiratory tract, CSF and sera of humans, bat feces, chimpanzee feces, muscle tissues of chickens, cows and goats, and insect abdomens (Dayaram *et al.*, 2013; Ge *et al.*, 2011; Li *et al.*, 2010a; Li *et al.*, 2011a; Li *et al.*, 2010b; Padilla-Rodriguez *et al.*, 2013; Phan *et al.*, 2014; Rosario *et al.*, 2011; Smits *et al.*, 2013b; Tan *et al.*, 2013). Consistent with the wide range of host animals, we found a high incidence of cycloviruses in shrews. Our phylogenetic analysis revealed that all identified sequences are closely related to cycloviruses, which were initially identified in CSF from human patients with CNS

manifestations (Smits *et al.*, 2013b; Tan *et al.*, 2013), raising the possibility of cross-species transmission between humans and shrews or rodents. Close sequence identity of cyclovirus species CyCV-VN has been observed between humans and domestic animals (Tan *et al.*, 2013). Although circovirus infection causes various clinical manifestations in birds and pigs, the pathogenicity of cycloviruses remains to be determined (Delwart & Li, 2012). Therefore, it is difficult to estimate the current risk of endemic cyclovirus in shrews and rodents.

The family *Picornaviridae* is a highly diverse virus family comprising 26 genera (Adams et al. 2013) (http://www.picornaviridae.com), and the continuous discovery of new species has further expanded the diversity of this family (Boros *et al.*, 2013; Boros *et al.*, 2012; Honkavuori *et al.*, 2011; Kapoor *et al.*, 2008a; Kapoor *et al.*, 2008b; Li *et al.*, 2009; Lim *et al.*, 2014; Ng *et al.*, 2012; Reuter *et al.*, 2012; Sauvage *et al.*, 2012; Woo *et al.*, 2012; Woo *et al.*, 2010). A number of picornaviruses have been detected in mammalian feces, but picornavirus has not been reported in shrews. Here, we identified the novel shrew picornavirus CroV1, which is distantly related to members of *Parechovirus*. Our findings broaden the current knowledge of genetic diversity of *Picornaviridae*. Unfortunately, the complete sequence of the 5′ UTR was unavailable; therefore, the IRES element in the CroV1 genome was not characterized.

CHV1, CHV2a and CHV2b have dicistronic genomes consisting of two nonoverlapping large ORFs encoding nonstructural and structural polyproteins. A similar genome organization has been observed in some picorna-like viruses, such as members of the family *Dicistroviridae*, the genera *Bacillarnavirus* and *Labyrnavirus*, and picalivirus A (Bonning & Miller, 2010; Ng *et al.*, 2012; Shirai *et al.*, 2006; Takao *et al.*, 2006). However, CHV1, CHV2a and CHV2b are phylogenetically quite distinct from these viruses and

318 picornaviruses. CHV1 is an unclassified picorna-like virus identified in human stool.
319 Nucleotide composition analysis suggested that CHV1 belongs to the insect host virus
320 group, and the detection of CHV1 in human stools was assumed to reflect
321 insect-contaminated food intake (Kapoor *et al.*, 2010). Interestingly, the closely related
322 viruses CHV2a and CHV2b were identified in the intestinal contents from shrews and
323 rodents. Given the insectivorous habit of shrews, these viruses might also reflect insect
324 consumption. Our nucleotide composition analysis also inferred an arthropod origin for
325 CHV2a and CHV2b. However, for rodents, the transmission route is difficult to estimate
326 and might be similar to the human case. A subsequent survey of calheviruses or related
327 viruses will provide insights into the distribution and host tropism of calheviruses.
328 In the present study, a combination strategy of viral nucleic acid enrichment and
329 subsequent high-throughput sequencing analysis revealed the enteric virome of *Crocidura*
330 spp. This initial description of the shrew enteric virome resulted in the discovery of novel
331 viruses. Subsequent analyses yielded complete or almost complete genome sequences of
332 these viruses and provided deep phylogenies. Consequently, these viruses can be
333 considered novel viral species. Our study provides an initial comprehensive view of the
334 diversity and distinctiveness of the shrew enteric virome, and also increases our
335 understanding of the viral diversity in mammals.
336

337 **Materials and Methods**

338 **Ethics Statement**

339 Samples were collected from wild shrews and rodents with permission from the Zambia
340 Wildlife Authority (Act No.12 of 1998). All rodents and shrews were euthanized by
341 inhalation of diethyl ether prior to dissection.

342

343 **Sample information**

344 We captured 24 shrews and 48 rodents around houses and fields using Sherman traps and
345 cage traps in Mpulungu, the northern province of Zambia, in 2012. After euthanasia, lung,
346 liver, spleen, kidney and intestinal contents were collected. Species were verified based on
347 the nucleotide sequence of the mitochondrial cytochrome *b* gene (Sasaki *et al.*, 2014).

348

349 **Enrichment and isolation of viral nucleic acids from shrew intestinal contents**

350 Viral nucleic acids were isolated and enriched for high-throughput sequencing as described
351 previously (Donaldson *et al.*, 2010; Phan *et al.*, 2011; Wu *et al.*, 2012) with some
352 modifications. In brief, aliquots of 700 µl of Hank's Balanced Salt Solution were added to
353 the intestinal contents of each shrew (100-200 mg). The suspensions were vortexed until
354 well-blended and were centrifuged at 10,000 × *g* for 3 min. Aliquots of 250 µl of each
355 clarified supernatant were pooled and filtered through a Minisart 0.45-µm syringe filter
356 (Sartorius) to remove unpelleted bacterial-size substances. The filtrate was concentrated
357 and buffer-exchanged into 800 µl of fresh Hank's Balanced Salt Solution using Amicon
358 Ultra-15 Centrifugal Filter Units with Ultracel-50 membranes (Merck Millipore). The
359 concentrated filtrate was treated with a cocktail of nuclease enzymes consisting of 10 µl of
360 TURBO DNase (20 U, Ambion; Life Technologies), 0.5 µl of benzonase (125 U,

Sigma-Aldrich) and 8 µl of 10 mg/ml RNase A (Roche Diagnostics) in 1× TURBO DNase buffer (Ambion) at 37 °C for 1 h to digest naked nucleic acids. Viral nucleic acids within viral capsids are resistant to nuclease digestion (Allander *et al.*, 2001). Then, 400 µl of the sample solution was processed using a High Pure Viral Nucleic Acid kit (Roche Diagnostics) to extract nucleic acids from DNA viruses according to the manufacturer's protocol, with the exception that 10 µg of linear polyacrylamide (Sigma) was used as the carrier instead of the carrier RNA supplied with the kit (Malboeuf *et al.*, 2013). The remaining 420 µl of the sample solution was processed using a QIAamp Viral RNA Mini kit to extract nucleic acids from RNA viruses (Qiagen) according to the manufacturer's protocol, with the exception that 15 µg of linear polyacrylamide was used as the carrier.

cDNA synthesis and sequence-independent amplification

Double-stranded cDNA was synthesized using the sequence-tagged random hexamer (5′-cgctcttccgatctNNNNNN-3′) (Yozwiak *et al.*, 2010) using the cDNA Synthesis kit (TAKARA BIO) according to the manufacturer's protocol and then purified using the Agencourt AMPure XP kit (Beckman Coulter). Sequence-independent amplification was performed with a tag sequence primer (5′-cgctcttccgatct-3′) and Ex Taq Hot Start Version (TAKARA BIO). The PCR cycling was performed as follows: 94 °C for 1 min, followed by 30 cycles of 98 °C for 10 s, 40 °C for 30 s and 72 °C for 1 min, with a final extension at 72 °C for 5 min.

Library preparation and high-throughput sequencing on the Ion-PGM system

Library preparation and high-throughput sequencing were performed according to the manufacturer's protocols provided by Ion Torrent (Life Technologies). In brief, the

385 extracted viral DNA sample and total amplified cDNA sample were pooled and sheared
386 using a Covaris S2 focused-ultrasonicator (Covaris) following the 400 bp protocol. From
387 this fragmented sample, a 400-base-read library was prepared using the Ion Plus Fragment
388 Library kit (Ion Torrent) and E-Gel SizeSelect 2% Agarose Gels (Invitrogen; Life
389 Technologies). Emulsion PCR was performed using the diluted library (13 pM) with the Ion
390 PGM Template OT2 400 kit (Ion Torrent). Sequencing was performed using the Ion PGM
391 Sequencing 400 kit, the Ion 318 Chip V2 and the Ion PGM sequencer (Ion Torrent). The
392 raw sequence data from the metagenomic analysis have been deposited in the Sequence
393 Read Archive of GenBank/EMBL/DDBJ (accession number DRA002561).
394
395 **Taxonomic assignment**
396 Unassembled sequence reads were compared with NCBI nucleotide database (nt) by using
397 BLASTN (version 2.2.26+). Results with an E-value ≤ 0.0001 were selected and used for
398 taxonomic classification by MEGAN (version 4.62.5) (Huson *et al.*, 2007). The lowest
399 common ancestor algorithm with parameters of minimum support = 5, minimum score = 25,
400 top percent = 10, and win score = 0 was used to compute the taxonomic content of the
401 sequences.
402
403 **Genome sequencing of novel cycloviruses**
404 The complete genome sequences of novel shrew and rodent cycloviruses were amplified
405 using nucleic acids from each individual shrew or rodent sample by inverse PCR with Tks
406 Gflex DNA polymerase (TAKARA BIO) and a primer set targeting *Rep*,
407 (5′-GAGTCCCTGTCAAAGGAGGATATGA-3′) and
408 (5′-TCKRTAAGGRTATCKGTCGCAGATCTTG-3′). Amplicons were purified using the

QIAquick Gel Extraction kit (Qiagen), cloned into the pCR4Blunt-TOPO vector (Invitrogen) and then sequenced by Sanger sequencing with primer walking. The sequence region recognized by the primer set targeting *Rep* was amplified by viral species-specific primers to confirm the true sequence.

Genome sequencing of novel linear viruses

To determine the genome sequence of CroV1, CHV2a and CHV2b in intestinal contents, the overlapping large fragments were amplified by RT-PCR using primers designed from the high-throughput sequencing reads. RACE was performed to obtain the 5′ and 3′ UTR sequences using the SMARTer RACE cDNA Amplification kit (Clontech) or an alternative strategy using the DT88 adaptor as reported previously (Li *et al.*, 2005). All amplified fragments were sequenced by Sanger sequencing with primer walking and assembled manually using GENETYX software ver. 10 (GENETYX).

Genetic characterization and phylogenetic analysis

Stem-loop structures of cycloviruses with the nonamer sequence were predicted by the Mfold webserver (Zuker, 2003). SimPlot sliding window analysis was performed by SimPlot software ver. 3.5.1 with a window size of 200 amino acids and a step size of 5 amino acids (Lole *et al.*, 1999). For phylogenetic analysis, reference sequences were obtained from the GenBank database, and multiple sequence alignments were constructed using the ClustalW and MEGA 6 packages (Tamura *et al.*, 2013; Thompson *et al.*, 1994). Bayesian phylogenetic analysis was performed using MrBayes software version 3.2.2 (Ronquist *et al.*, 2012) with the WAG amino acid substitution model. The obtained trees were visualized with FigTree software, version 1.4.

433

434 **Nucleotide composition analysis and canonical discriminant analysis**

435 Nucleotide composition analysis was performed as described previously (Kapoor *et al.*,
436 2010). Mononucleotide and dinucleotide frequencies for each viral sequence were obtained
437 using the composition scan program in the SSE package (Simmonds, 2012). Canonical
438 discriminant analysis was performed using the RAFisher2cda program (Trujillo-Ortiz *et al.*,
439 2004). The genome sequences of 112 vertebrate-derived viruses, 64 arthropod-derived
440 viruses and 171 plant-derived viruses classified as picorna-like viruses that were used as
441 reference sequences for the analysis are listed in Shan *et al*., 2011.

442

443 **PCR/RT-PCR screening**

444 To screen for the identified viruses, DNA and RNA were extracted from intestinal contents
445 suspensions using the High Pure Viral Nucleic Acid kit and High Pure Viral RNA kit
446 (Roche Diagnostics), respectively. Tissue DNA and RNA were extracted using the QIAamp
447 DNA Mini kit (Qiagen), a combination of TRIzol reagent and the PureLink RNA Mini kit
448 (Ambion), or the AllPrep DNA/RNA Mini kit (Qiagen). PCR screening for novel
449 cycloviruses was performed using the Tks Gflex DNA polymerase. The PCR cycling was
450 performed as follows: 94 °C for 1 min, followed by 35 cycles of 98 °C for 10 s, 65 °C for
451 15 s and 68 °C for 1 min, with a final extension at 68 °C for 5 min. RT-PCR screening for
452 CroV1, CHV2a and CHV2b was performed using the SuperScript III One-Step RT-PCR
453 System with Platinum Taq (Invitrogen). The one-step RT-PCR cycling was performed as
454 follows: 60 °C for 1 min, 50 °C for 30 min, and 94 °C for 2 min, followed by 40 cycles of
455 94 °C for 15 s, 56 °C for 30 s and 68 °C for 1 min, with a final extension at 68 °C for 5 min.
456 The following primers were used in this screening experiment: (5′-

GAGTCCCTGTCAAAGGAGGATATGA -3′) and (5′-TCKRTAAGGRTATCKGTCGCAGATCTTG -3′) for cyclovirus screening; (5′-CACACTGGAATATCGATTGAGGAAG -3′) and (5′-CAACACAGTTGTACAAGGAGATCCA -3′) for CroV1 screening; (5′-GATTGCTGCGTTTAAGTCGCTAGA -3′) and (5′-AAATCGCCGCTTGAGAAACGTGA -3′) for CHV2a screening; and (5′-CTCGGATGTCTTTGGAAGTGACTG -3′) and (5′-AAGCTGCGTGTACACTTCCTCAAG -3′) for CHV2b screening. All positive PCR/RT-PCR signals were confirmed by direct sequencing.

Acknowledgments

We thank the Zambia Wildlife Authority for their support of this research. This study was supported by the Japan Initiative for Global Research Network of Infectious Diseases (J-GRID) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan, and Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 24405043.

473 **References**

474

475 **Adams, M. J., King, A. M. Q. & Carstens, E. B. (2013).** Ratification vote on taxonomic
476 proposals to the International Committee on Taxonomy of Viruses (2013). Arch Virol
477 158, 2023–2030.

478 **Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. (2001).** A virus
479 discovery method incorporating DNase treatment and its application to the
480 identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* **98**,
481 11609-11614.

482 **Anthony, S. J., Epstein, J. H., Murray, K. A., Navarrete-Macias, I.,**
483 **Zambrana-Torrelio, C. M., Solovyov, A., Ojeda-Flores, R., Arrigo, N. C., Islam,**
484 **A., Ali Khan, S., Hosseini, P., Bogich, T. L., Olival, K. J., Sanchez-Leon, M. D.,**
485 **Karesh, W. B., Goldstein, T., Luby, S. P., Morse, S. S., Mazet, J. A., Daszak, P.**
486 **& Lipkin, W. I. (2013).** A strategy to estimate unknown viral diversity in mammals.
487 *MBio* **4**, e00598-00513.

488 **Baker, K. S., Leggett, R. M., Bexfield, N. H., Alston, M., Daly, G., Todd, S., Tachedjian,**
489 **M., Holmes, C. E., Crameri, S., Wang, L. F., Heeney, J. L., Suu-Ire, R., Kellam,**
490 **P., Cunningham, A. A., Wood, J. L., Caccamo, M. & Murcia, P. R. (2013).**
491 Metagenomic study of the viruses of African straw-coloured fruit bats: detection of
492 a chiropteran poxvirus and isolation of a novel adenovirus. *Virology* **441**, 95-106.

493 **Blom, N., Hansen, J., Blaas, D. & Brunak, S. (1996).** Cleavage site analysis in
494 picornaviral polyproteins: discovering cellular targets by neural networks. *Protein*
495 *Sci* **5**, 2203-2216.

496 **Bodewes, R., van der Giessen, J., Haagmans, B. L., Osterhaus, A. D. & Smits, S. L.**

(2013). Identification of multiple novel viruses, including a parvovirus and a hepevirus, in feces of red foxes. *J Virol* **87**, 7758-7764.

Bonning, B. C. & Miller, W. A. (2010). Dicistroviruses. *Annu Rev Entomol* **55**, 129-150.

Boros, A., Nemes, C., Pankovics, P., Kapusinszky, B., Delwart, E. & Reuter, G. (2013). Genetic characterization of a novel picornavirus in turkeys (Meleagris gallopavo) distinct from turkey galliviruses and megriviruses and distantly related to the members of the genus Avihepatovirus. *J Gen Virol* **94**, 1496-1509.

Boros, Á., Nemes, C., Pankovics, P., Kapusinszky, B., Delwart, E. & Reuter, G. (2012). Identification and complete genome characterization of a novel picornavirus in turkey (Meleagris gallopavo). *J Gen Virol* **93**, 2171-2182.

Dacheux, L., Cervantes-Gonzalez, M., Guigon, G., Thiberge, J. M., Vandenbogaert, M., Maufrais, C., Caro, V. & Bourhy, H. (2014). A preliminary study of viral metagenomics of French bat species in contact with humans: identification of new mammalian viruses. *PLoS One* **9**, e87194.

Dayaram, A., Potter, K. A., Moline, A. B., Rosenstein, D. D., Marinov, M., Thomas, J. E., Breitbart, M., Rosario, K., Argüello-Astorga, G. R. & Varsani, A. (2013). High global diversity of cycloviruses amongst dragonflies. *J Gen Virol* **94**, 1827-1840.

Delwart, E. (2013). A roadmap to the human virome. *PLoS Pathog* **9**, e1003146.

Delwart, E. & Li, L. (2012). Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other Rep encoding small circular ssDNA genomes. *Virus Res* **164**, 114-121.

Donaldson, E. F., Haskew, A. N., Gates, J. E., Huynh, J., Moore, C. J. & Frieman, M. B. (2010). Metagenomic analysis of the viromes of three North American bat

species: viral diversity among different bat species that share a common habitat. *J Virol* **84**, 13004-13018.

Dubey, S., Antonin, M., Denys, C. & Vogel, P. (2007). Use of phylogeny to resolve the taxonomy of the widespread and highly polymorphic African giant shrews (Crocidura olivieri group, Crocidurinae, Mammalia). *Zoology (Jena)* **110**, 48-57.

Dürrwald, R., Kolodziejek, J., Weissenböck, H. & Nowotny, N. (2014). The bicolored white-toothed shrew Crocidura leucodon (HERMANN 1780) is an indigenous host of mammalian Borna disease virus. *PLoS One* **9**, e93659.

Esposito, S., Rahamat-Langendoen, J., Ascolese, B., Senatore, L., Castellazzi, L. & Niesters, H. G. (2014). Pediatric parechovirus infections. *J Clin Virol* **60**, 84-89.

Firth, C. & Lipkin, W. I. (2013). The genomics of emerging pathogens. *Annu Rev Genomics Hum Genet* **14**, 281-300.

Ge, X., Li, J., Peng, C., Wu, L., Yang, X., Wu, Y., Zhang, Y. & Shi, Z. (2011). Genetic diversity of novel circular ssDNA viruses in bats in China. *J Gen Virol* **92**, 2646-2653.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y. & Shi, Z. (2012). Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* **86**, 4620-4630.

Goodfellow, I. (2011). The genome-linked protein VPg of vertebrate viruses - a multifaceted protein. *Curr Opin Virol* **1**, 355-362.

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

Guo, W. P., Lin, X. D., Wang, W., Tian, J. H., Cong, M. L., Zhang, H. L., Wang, M. R., Zhou, R. H., Wang, J. B., Li, M. H., Xu, J., Holmes, E. C. & Zhang, Y. Z. (2013). Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents. *PLoS Pathog* **9**, e1003159.

Handley, S. A., Thackray, L. B., Zhao, G., Presti, R., Miller, A. D., Droit, L., Abbink, P., Maxfield, L. F., Kambal, A., Duan, E., Stanley, K., Kramer, J., Macri, S. C., Permar, S. R., Schmitz, J. E., Mansfield, K., Brenchley, J. M., Veazey, R. S., Stappenbeck, T. S., Wang, D., Barouch, D. H. & Virgin, H. W. (2012). Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* **151**, 253-266.

Hilbe, M., Herrsche, R., Kolodziejek, J., Nowotny, N., Zlinszky, K. & Ehrensperger, F. (2006). Shrews as reservoir hosts of borna disease virus. *Emerg Infect Dis* **12**, 675-677.

Honkavuori, K. S., Shivaprasad, H. L., Briese, T., Street, C., Hirschberg, D. L., Hutchison, S. K. & Lipkin, W. I. (2011). Novel picornavirus in Turkey poults with hepatitis, California, USA. *Emerg Infect Dis* **17**, 480-487.

Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res* **17**, 377-386.

Joffret, M. L., Bouchier, C., Grandadam, M., Zeller, H., Maufrais, C., Bourhy, H., Despres, P., Delpeyroux, F. & Dacheux, L. (2013). Genomic characterization of Sebokele virus 1 (SEBV1) reveals a new candidate species among the genus Parechovirus. *J Gen Virol* **94**, 1547-1553.

Kamer, G. & Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res* **12**,

7269-7282.

Kapoor, A., Simmonds, P., Lipkin, W. I., Zaidi, S. & Delwart, E. (2010). Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *J Virol* **84**, 10322-10328.

Kapoor, A., Victoria, J., Simmonds, P., Slikas, E., Chieochansin, T., Naeem, A., Shaukat, S., Sharif, S., Alam, M. M., Angez, M., Wang, C., Shafer, R. W., Zaidi, S. & Delwart, E. (2008a). A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proc Natl Acad Sci U S A* **105**, 20482-20487.

Kapoor, A., Victoria, J., Simmonds, P., Wang, C., Shafer, R. W., Nims, R., Nielsen, O. & Delwart, E. (2008b). A highly divergent picornavirus in a marine mammal. *J Virol* **82**, 311-320.

King, A. M., Adams, M. J., Lefkowitz, E. J. & Carstens, E. B. (2012). *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*: Elsevier.

Le Gall, O., Christian, P., Fauquet, C. M., King, A. M., Knowles, N. J., Nakashima, N., Stanway, G. & Gorbalenya, A. E. (2008). Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. *Arch Virol* **153**, 715-727.

Li, L., Kapoor, A., Slikas, B., Bamidele, O. S., Wang, C., Shaukat, S., Masroor, M. A., Wilson, M. L., Ndjango, J. B., Peeters, M., Gross-Camp, N. D., Muller, M. N., Hahn, B. H., Wolfe, N. D., Triki, H., Bartkus, J., Zaidi, S. Z. & Delwart, E. (2010a). Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol* **84**, 1674-1682.

Li, L., Shan, T., Soji, O. B., Alam, M. M., Kunz, T. H., Zaidi, S. Z. & Delwart, E. (2011a). Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *J Gen Virol* **92**, 768-772.

Li, L., Shan, T., Wang, C., Côté, C., Kolman, J., Onions, D., Gulland, F. M. & Delwart, E. (2011b). The fecal viral flora of California sea lions. *J Virol* **85**, 9909-9917.

Li, L., Victoria, J., Kapoor, A., Blinkova, O., Wang, C., Babrzadeh, F., Mason, C. J., Pandey, P., Triki, H., Bahri, O., Oderinde, B. S., Baba, M. M., Bukbuk, D. N., Besser, J. M., Bartkus, J. M. & Delwart, E. L. (2009). A novel picornavirus associated with gastroenteritis. *J Virol* **83**, 12002-12006.

Li, L., Victoria, J. G., Wang, C., Jones, M., Fellers, G. M., Kunz, T. H. & Delwart, E. (2010b). Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol* **84**, 6955-6965.

Li, Z., Yu, M., Zhang, H., Wang, H. Y. & Wang, L. F. (2005). Improved rapid amplification of cDNA ends (RACE) for mapping both the 5' and 3' terminal sequences of paramyxovirus genomes. *J Virol Methods* **130**, 154-156.

Lim, E. S., Cao, S., Holtz, L. R., Antonio, M., Stine, O. C. & Wang, D. (2014). Discovery of rosavirus 2, a novel variant of a rodent-associated picornavirus, in children from The Gambia. *Virology* **454-455**, 25-33.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160.

Luke, G. A., de Felipe, P., Lukashev, A., Kallioinen, S. E., Bruno, E. A. & Ryan, M. D. (2008). Occurrence, function and evolutionary origins of '2A-like' sequences in

617 virus genomes. *J Gen Virol* **89**, 1036-1042.
618 **Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K.,**
619 **Lindberg, A. M., Persson, B., Allander, T. & Andersson, B. (2012).**
620 Characterization of the viral microbiome in patients with severe lower respiratory
621 tract infections, using metagenomic sequencing. *PLoS One* **7**, e30875.
622 **Malboeuf, C. M., Yang, X., Charlebois, P., Qu, J., Berlin, A. M., Casali, M., Pesko, K.**
623 **N., Boutwell, C. L., DeVincenzo, J. P., Ebel, G. D., Allen, T. M., Zody, M. C.,**
624 **Henn, M. R. & Levin, J. Z. (2013).** Complete viral RNA genome sequencing of
625 ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res*
626 **41**, e13.
627 **Meerburg, B. G., Singleton, G. R. & Kijlstra, A. (2009).** Rodent-borne diseases and their
628 risks for public health. *Crit Rev Microbiol* **35**, 221-270.
629 **Mokili, J. L., Rohwer, F. & Dutilh, B. E. (2012).** Metagenomics and future perspectives in
630 virus discovery. *Curr Opin Virol* **2**, 63-77.
631 **Morse, S. S., Mazet, J. A., Woolhouse, M., Parrish, C. R., Carroll, D., Karesh, W. B.,**
632 **Zambrana-Torrelio, C., Lipkin, W. I. & Daszak, P. (2012).** Prediction and
633 prevention of the next pandemic zoonosis. *Lancet* **380**, 1956-1965.
634 **Nakashima, N. & Uchiumi, T. (2009).** Functional analysis of structural motifs in
635 dicistroviruses. *Virus Res* **139**, 137-147.
636 **Ng, T. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L.,**
637 **Oderinde, B. S., Wommack, K. E. & Delwart, E. (2012).** High variety of known
638 and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* **86**,
639 12161-12175.
640 **Ng, T. F., Mesquita, J. R., Nascimento, M. S., Kondov, N. O., Wong, W., Reuter, G.,**

641 **Knowles, N. J., Vega, E., Esona, M. D., Deng, X., Vinjé, J. & Delwart, E. (2014).**
642 Feline fecal virome reveals novel and prevalent enteric viruses. *Vet Microbiol* **171**,
643 102-111.
644 **Niklasson, B., Kinnunen, L., Hörnfeldt, B., Hörling, J., Benemar, C., Hedlund, K. O.,**
645 **Matskova, L., Hyypiä, T. & Winberg, G. (1999).** A new picornavirus isolated from
646 bank voles (Clethrionomys glareolus). *Virology* **255**, 86-93.
647 **Padilla-Rodriguez, M., Rosario, K. & Breitbart, M. (2013).** Novel cyclovirus discovered
648 in the Florida woods cockroach Eurycotis floridana (Walker). *Arch Virol* **158**,
649 1389-1392.
650 **Phan, T. G., Kapusinszky, B., Wang, C., Rose, R. K., Lipton, H. L. & Delwart, E. L.**
651 **(2011).** The fecal viral flora of wild rodents. *PLoS Pathog* **7**, e1002218.
652 **Phan, T. G., Luchsinger, V., Avendaño, L. F., Deng, X. & Delwart, E. (2014).** Cyclovirus
653 in nasopharyngeal aspirates of Chilean children with respiratory infections. *J Gen*
654 *Virol* **95**, 922-927.
655 **Phan, T. G., Vo, N. P., Boros, Á., Pankovics, P., Reuter, G., Li, O. T., Wang, C., Deng, X.,**
656 **Poon, L. L. & Delwart, E. (2013a).** The viruses of wild pigeon droppings. *PLoS*
657 *One* **8**, e72787.
658 **Phan, T. G., Vo, N. P., Simmonds, P., Samayoa, E., Naccache, S., Chiu, C. Y. & Delwart,**
659 **E. (2013b).** Rosavirus: the prototype of a proposed new genus of the Picornaviridae
660 family. *Virus Genes* **47**, 556-558.
661 **Reuter, G., Pankovics, P., Knowles, N. J. & Boros, Á. (2012).** Two closely related novel
662 picornaviruses in cattle and sheep in Hungary from 2008 to 2009, proposed as
663 members of a new genus in the family Picornaviridae. *J Virol* **86**, 13295-13302.
664 **Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S.,**

Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539-542.

Rosario, K., Duffy, S. & Breitbart, M. (2012). A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* **157**, 1851-1871.

Rosario, K., Marinov, M., Stainton, D., Kraberger, S., Wiltshire, E. J., Collings, D. A., Walters, M., Martin, D. P., Breitbart, M. & Varsani, A. (2011). Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). *J Gen Virol* **92**, 1302-1308.

Sasaki, M., Muleya, W., Ishii, A., Orba, Y., Hang'ombe, B. M., Mweene, A. S., Moonga, L., Thomas, Y., Kimura, T. & Sawa, H. (2014). Molecular epidemiology of paramyxoviruses in Zambian wild rodents and shrews. *J Gen Virol* **95**, 325-330.

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. *J Virol* **86**, 10036-10046.

Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A. & Delwart, E. (2011). The fecal virome of pigs on a high-density farm. *J Virol* **85**, 11697-11708.

Shirai, Y., Takao, Y., Mizumoto, H., Tomaru, Y., Honda, D. & Nagasaki, K. (2006). Genomic and phylogenetic analysis of a single-stranded RNA virus infecting Rhizosolenia setigera (Stramenopiles: Bacillariophyceae). *J Mar Biol Ass U K* **86**, 475-483.

Simmonds, P. (2012). SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* **5**, 50.

689 **Smith, I. & Wang, L. F. (2013).** Bats and their virome: an important source of emerging
690 viruses capable of infecting humans. *Curr Opin Virol* **3**, 84-91.
691 **Smits, S. L., Raj, V. S., Oduber, M. D., Schapendonk, C. M., Bodewes, R., Provacia, L.,**
692 **Stittelaar, K. J., Osterhaus, A. D. & Haagmans, B. L. (2013a).** Metagenomic
693 analysis of the ferret fecal viral flora. *PLoS One* **8**, e71595.
694 **Smits, S. L., Zijlstra, E. E., van Hellemond, J. J., Schapendonk, C. M., Bodewes, R.,**
695 **Schürch, A. C., Haagmans, B. L. & Osterhaus, A. D. (2013b).** Novel cyclovirus
696 in human cerebrospinal fluid, Malawi, 2010-2011. *Emerg Infect Dis* **19**.
697 **Sweeney, T. R., Cisnetto, V., Bose, D., Bailey, M., Wilson, J. R., Zhang, X., Belsham, G.**
698 **J. & Curry, S. (2010).** Foot-and-mouth disease virus 2C is a hexameric AAA+
699 protein with a coordinated ATP hydrolysis mechanism. *J Biol Chem* **285**,
700 24347-24359.
701 **Takao, Y., Mise, K., Nagasaki, K., Okuno, T. & Honda, D. (2006).** Complete nucleotide
702 sequence and genome organization of a single-stranded RNA virus infecting the
703 marine fungoid protist Schizochytrium sp. *J Gen Virol* **87**, 723-733.
704 **Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013).** MEGA6:
705 Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**,
706 2725-2729.
707 **Tan, l. V., van Doorn, H. R., Nghia, H. D., Chau, T. T., Tu, l. T., de Vries, M., Canuti,**
708 **M., Deijs, M., Jebbink, M. F., Baker, S., Bryant, J. E., Tham, N. T., BKrong, N.**
709 **T., Boni, M. F., Loi, T. Q., Phuong, l. T., Verhoeven, J. T., Crusat, M., Jeeninga,**
710 **R. E., Schultsz, C., Chau, N. V., Hien, T. T., van der Hoek, L., Farrar, J. & de**
711 **Jong, M. D. (2013).** Identification of a new cyclovirus in cerebrospinal fluid of
712 patients with acute central nervous system infections. *MBio* **4**, e00231-00213.

713 **Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the
714 sensitivity of progressive multiple sequence alignment through sequence weighting,
715 position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**,
716 4673-4680.
717 **Trujillo-Ortiz, A., Hernandez-Walls, R. & Perez-Osuna, S. (2004).** RAFisher2cda:
718 canonical discriminant analysis. *A Matlab file*.
719 **van den Brand, J. M., van Leeuwen, M., Schapendonk, C. M., Simon, J. H.,**
720 **Haagmans, B. L., Osterhaus, A. D. & Smits, S. L. (2012).** Metagenomic analysis
721 of the viral flora of pine marten and European badger feces. *J Virol* **86**, 2360-2365.
722 **Williams, C. H., Panayiotou, M., Girling, G. D., Peard, C. I., Oikarinen, S., Hyöty, H.**
723 **& Stanway, G. (2009).** Evolution and conservation in human parechovirus genomes.
724 *J Gen Virol* **90**, 1702-1712.
725 **Wilson, D. E. & Reeder, D. M. (2011).** Animal biodiversity: An outline of higher-level
726 classification and survey of taxonomic richness. *Zootaxa* **3148**, 56-60.
727 **Witkowski, P. T., Klempa, B., Ithete, N. L., Auste, B., Mfune, J. K., Hoveka, J.,**
728 **Matthee, S., Preiser, W. & Kruger, D. H. (2014).** Hantaviruses in Africa. *Virus Res*
729 **187**, 34-42.
730 **Woo, P. C., Lau, S. K., Choi, G. K., Huang, Y., Teng, J. L., Tsoi, H. W., Tse, H., Yeung,**
731 **M. L., Chan, K. H., Jin, D. Y. & Yuen, K. Y. (2012).** Natural occurrence and
732 characterization of two internal ribosome entry site elements in a novel virus, canine
733 picodicistrovirus, in the picornavirus-like superfamily. *J Virol* **86**, 2797-2808.
734 **Woo, P. C., Lau, S. K., Huang, Y., Lam, C. S., Poon, R. W., Tsoi, H. W., Lee, P., Tse, H.,**
735 **Chan, A. S., Luk, G., Chan, K. H. & Yuen, K. Y. (2010).** Comparative analysis of
736 six genome sequences of three novel picornaviruses, turdiviruses 1, 2 and 3, in dead

737 wild birds, and proposal of two novel genera, Orthoturdivirus and Paraturdivirus, in
738 the family Picornaviridae. *J Gen Virol* **91**, 2433-2448.
739 **Wu, Z., Ren, X., Yang, L., Hu, Y., Yang, J., He, G., Zhang, J., Dong, J., Sun, L., Du, J.,**
740 **Liu, L., Xue, Y., Wang, J., Yang, F., Zhang, S. & Jin, Q. (2012).** Virome analysis
741 for identification of novel mammalian viruses in bat species from Chinese provinces.
742 *J Virol* **86**, 10999-11012.
743 **Yanagihara, R., Gu, S. H., Arai, S., Kang, H. J. & Song, J. W. (2014).** Hantaviruses:
744 Rediscovery and new beginnings. *Virus Res* **187**, 6-14.
745 **Yozwiak, N. L., Skewes-Cox, P., Gordon, A., Saborio, S., Kuan, G., Balmaseda, A.,**
746 **Ganem, D., Harris, E. & DeRisi, J. L. (2010).** Human enterovirus 109: a novel
747 interspecies recombinant enterovirus isolated from a case of acute pediatric
748 respiratory illness in Nicaragua. *J Virol* **84**, 9047-9058.
749 **Yu, J. M., Li, X. Y., Ao, Y. Y., Li, L. L., Liu, N., Li, J. S. & Duan, Z. J. (2013).**
750 Identification of a novel picornavirus in healthy piglets and seroepidemiological
751 evidence of its presence in humans. *PLoS One* **8**, e70137.
752 **Zuker, M. (2003).** Mfold web server for nucleic acid folding and hybridization prediction.
753 *Nucleic Acids Res* **31**, 3406-3415.
754
755

Figure Legends

Fig. 1. Taxonomic classification of sequence reads from shrew intestinal contents.

The proportions of whole-sequence reads (A), viral-sequence reads (B) and the type of host predicted to be associated with the virus from which the sequence reads were derived (C) are shown in the charts. The numbers in parentheses indicate the percentage of sequence reads related to members of each taxon. dsDNA, double-stranded DNA; dsRNA, double-stranded RNA; ssDNA, single-stranded DNA; ssRNA, single-stranded RNA.

Fig. 2. Genome organization and phylogenetic relationship of the cycloviruses identified.

(A) Diagrams of the predicted genome organization of the identified cycloviruses. Black arrows indicate ORFs encoding rolling circle replication initiator protein (Rep). White arrows indicate ORFs encoding capsid protein (Cap). Gray arrows indicate hypothetical ORFs with unknown functions. The positions of the conserved rolling circle amplification (RCA) motifs RCA I (WTLNN), RCA II (HLQGFCNL) and RCA III (YCSKGGD) and the helicase motifs Walker A (GCTGTGKS), B (VVIDDFYGW) and C (ITSE) are indicated by black and white arrowheads, respectively. (B) Predicted stem-loop structure of cyclovirus CyCV/ZM01. The highly conserved nonamer sequence is highlighted in grey. (C) Phylogenetic analysis of the full-length Rep of representative cycloviruses and cycloviruses identified in this study. The respective accession numbers of the viral sequences are shown in parentheses. Bayesian posterior probabilities are indicated at each tree root. The scale bar represents a distance of 0.2 substitutions per site.

Fig. 3. Genome organization and phylogenetic relationships of crohivirus 1 (CroV1).

(A) Diagram of the predicted genome organization of CroV1. The P1 region consists of structural proteins. The P2 and P3 regions consist of nonstructural proteins. The positions of the cleavage sites in the polyprotein are indicated by white arrowheads with the nucleotide numbers. The characteristic motifs mapped in the diagram are as follows: the parechovirus-conserved motif (KxKxxRxK), the ribosomal skipping 2A motif (DxExNPGP), the helicase motifs Walker A (GxxGxGKS) and Walker B (DD), the picornavirus-conserved tyrosine residue (Y), the protease catalytic triad residues (H-D-C), the protease active motifs (GxCG and GxH) and the 3D polymerase motifs (KDELR, GxPSG, YGDD and FLKR). (B) SimPlot sliding window analysis of CroV1 compared with human parechovirus 1 (red line), Ljungan virus 87-012 (green line) and Sebokele virus 1 (blue line). A window size of 200 amino acids and a step size of 5 amino acids were used. (C) Phylogenetic analysis of the full-length 3D polymerase of representative picornaviruses and CroV1. The accession numbers of the picornavirus sequences are shown in parentheses. Bayesian posterior probabilities are indicated at each tree root. The scale bar represents a distance of 0.2 substitutions per site.

Fig. 4. Genome organization and phylogenetic relationships of the calheviruses identified.

(A) Diagram of the predicted genome organization of calhevirus 2a. Black and white boxes show the putative nonstructural polyprotein and structural polyprotein, respectively. The gray box shows a hypothetical ORF with unknown function. The positions of the helicase Walker A (GxxGxGKS) and B motifs (DD), the protease catalytic triad residues (H-D-S) and the highly conserved RNA-dependent RNA polymerase motifs (KDELR, YGDD and FLKR) are shown. (B) Phylogenetic analysis of the predicted RNA-dependent RNA

804 polymerase-encoding region of calhevirus 2a, calhevirus 2b, picorna-like viruses,
805 picornaviruses, and caliciviruses. The accession numbers of the viral sequences are shown
806 in parentheses. Bayesian posterior probabilities are indicated at each tree root. The scale bar
807 represents a distance of 0.2 substitutions per site.
808

Table 1 - PCR/RT-PCR screening results

The results are presented as the number of PCR or RT-PCR-positive individuals per number of shrews or rodents tested.

		CyCVs	CroV1	CHV2a	CHV2b
Shrew samples					
	Intestinal content	21/23	3/23	1/23	4/23
	Lung	0/8	0/3	0/1	0/4
	Liver	0/8	3/3	0/1	0/4
	Spleen	0/8	1/3	0/1	0/4
	Kidney	0/8	0/3	0/1	0/4
Rodent samples					
	Intestinal content	18/48	0/48	0/48	3/48
	Lung	0/8	-	-	0/3
	Liver	0/8	-	-	0/3
	Spleen	0/8	-	-	0/3
	Kidney	0/8	-	-	0/3

814 **Table 2** - Pairwise genomic nucleotide sequence identities between different cycloviruses.

Isolates	Percentage of nucleotide sequence identity					
	ZM01	**ZM41**	**ZM36a**	**ZM62**	**ZM50a**	**ZM32**
CyCV/ZM01 (AB937981)		81.4	76.2	78.4	78.1	78.9
CyCV/ZM41 (AB937984)	81.4		75.8	78.0	79.2	79.0
CyCV/ZM36a (AB937982)	76.2	75.8		73.9	77.1	77.8
CyCV/ZM62 (AB937987)	78.4	78.0	73.9		79.0	79.8
CyCV/ZM50a (AB937985)	78.1	79.2	77.1	79.0		91.9
CyCV/ZM32 (AB937980)	78.9	79.0	77.8	79.8	91.9	
CyCV/CyCV-VN (KF031465)	78.5	78.5	74.1	81.6	81.0	81.9
CyCV/VS5700009 (NC_021568)	64.0	66.3	67.5	64.7	63.4	63.9
CyCV/TN18 (GQ404858)	67.0	67.7	67.1	68.7	69.0	68.9

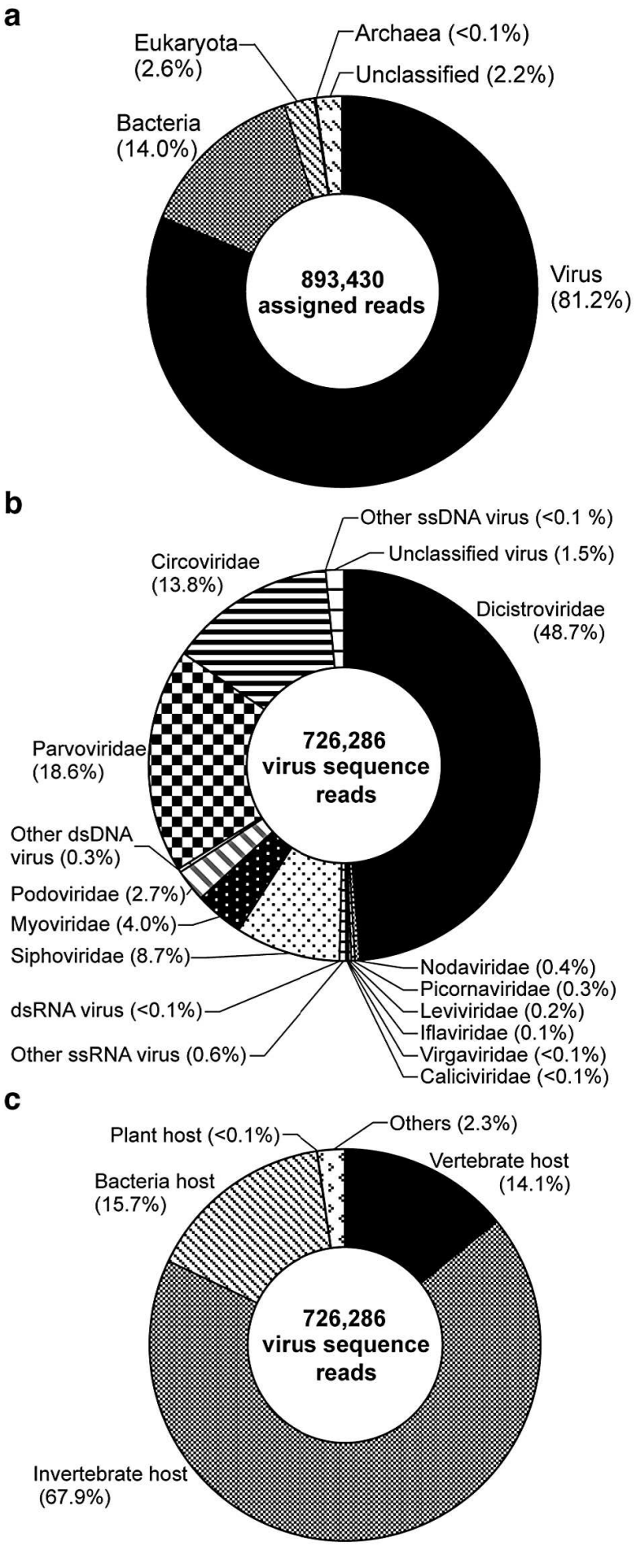
815

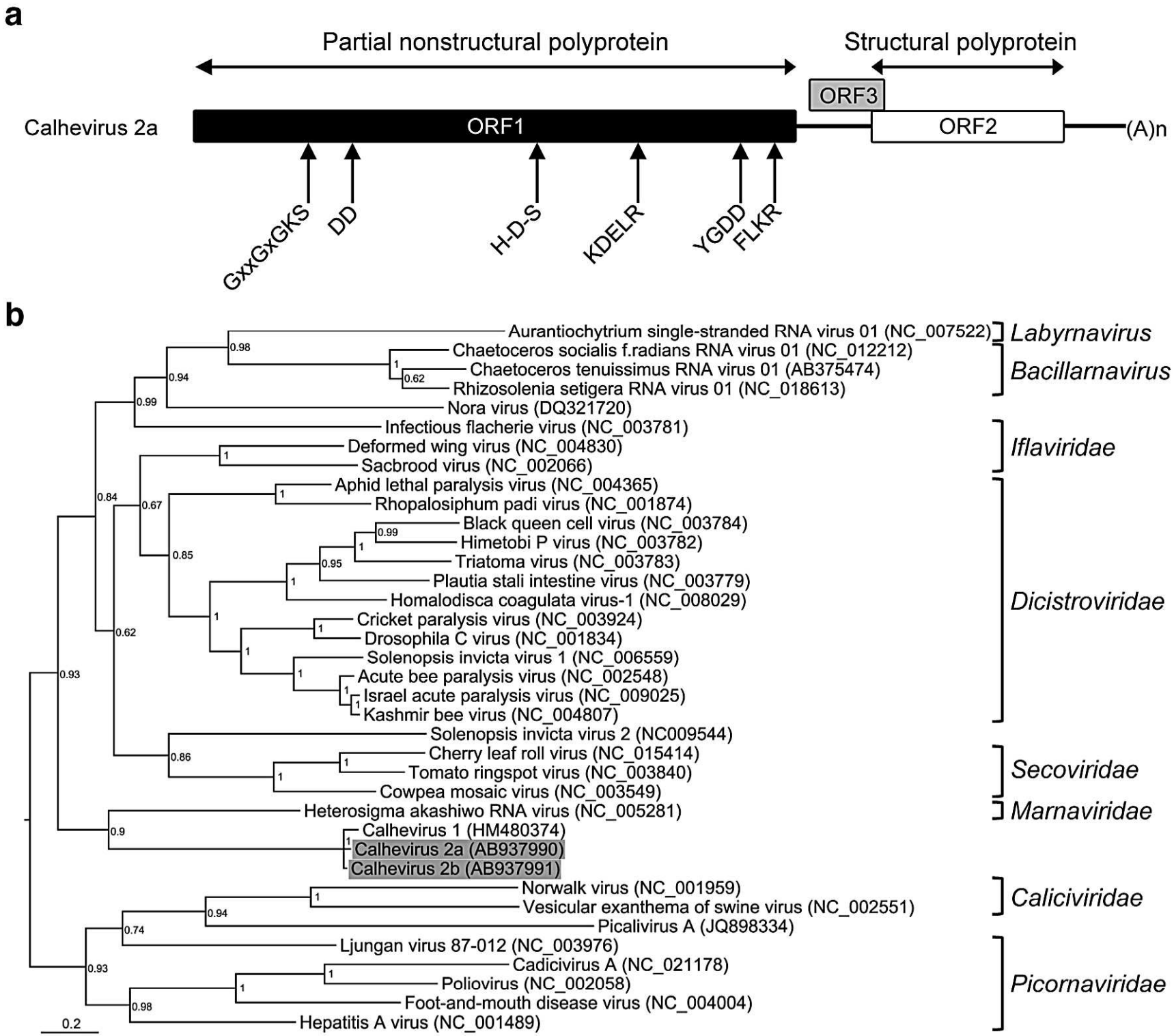
816

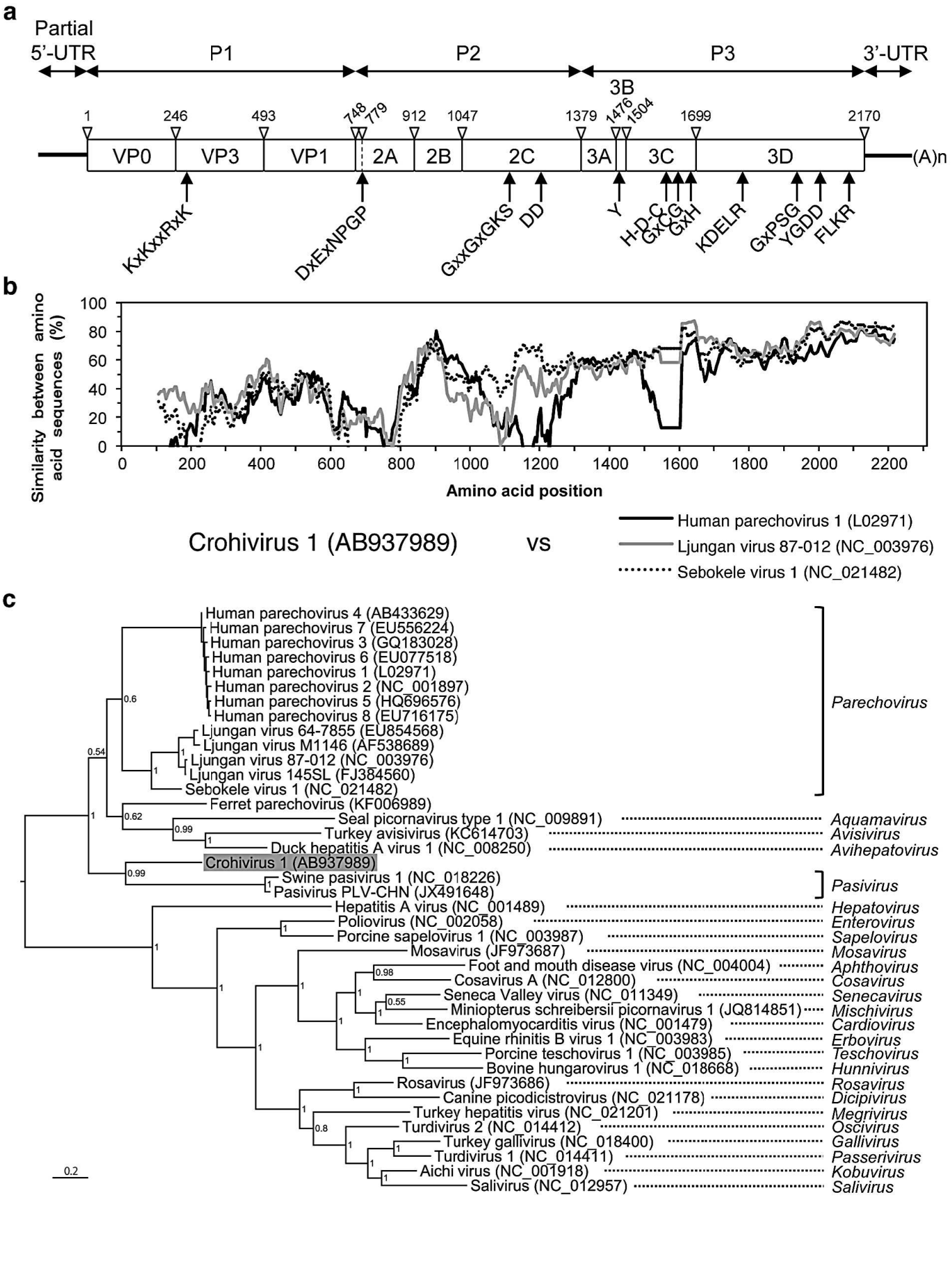
817 **Table 3** - Pairwise amino acid identities in the P1, P2 and P3 regions between crohivirus 1
818 and related members of the family *Picornavirus*

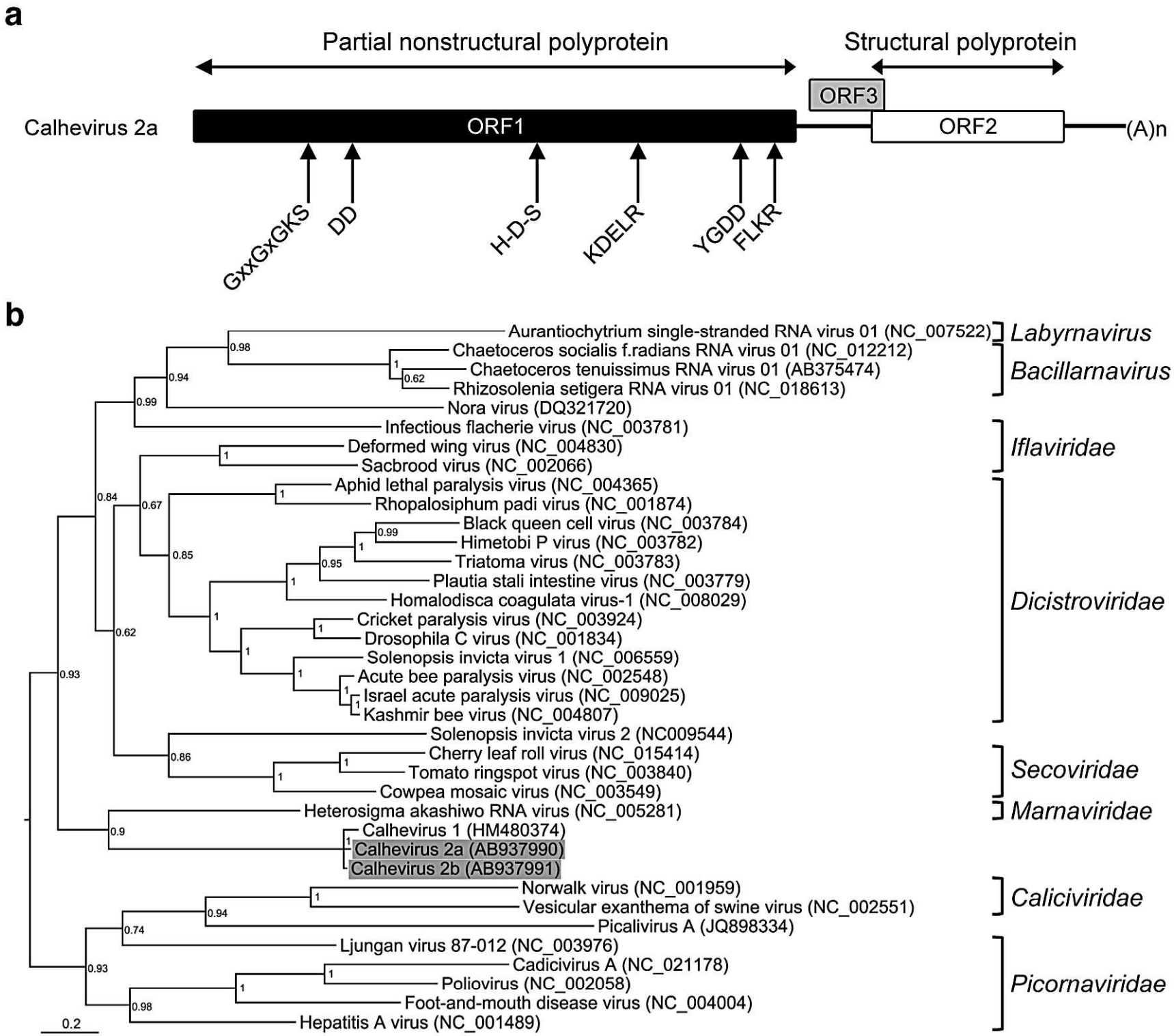
Genus	Species	Amino acid identities (%) with Crohivirus 1		
		P1	P2	P3
Parechovirus	Ljungan virus 64-7855 (EU854568)	27.2	37.8	38.8
	Ljungan virus 87-012 (NC_003976)	31.0	37.2	39.0
	Ljungan virus M1146 (AF538689)	27.2	36.8	39.7
	Ljungan virus 145SL (FJ384560)	31.3	37.1	38.9
	Human parechovirus 1 (EF051629)	33.5	28.1	35.7
	Human parechovirus 2 (NC_001897)	31.7	26.5	36.1
	Human parechovirus 3 (GQ183028)	33.9	27.3	35.9
	Human parechovirus 4 (AB433629)	27.8	28.1	35.7
	Human parechovirus 5 (HQ696576)	26.9	26.5	35.2
	Human parechovirus 6 (EU077518)	32.6	27.3	35.7
	Human parechovirus 7 (EU556224)	31.9	27.2	35.3
	Human parechovirus 8 (EU716175)	31.6	27.9	35.6
	Sebokele virus 1 (NC_021482)	31.4	38.2	37.3
Parechovirus-related	Ferret parechovirus (KF006989)	33.8	29.5	34.4
Pasivirus	Swine pasivirus 1 (NC_018226)	32.4	21.6	32.5
Avihepatovirus	Duck hepatitis A virus 1 (NC_008250)	19.0	26.0	30.3
Aquamavirus	Seal picornavirus type 1 (NC_009891)	12.4	13.1	24.1

819


a
Eukaryota (2.6%)
Archaea (<0.1%)
Unclassified (2.2%)
Bacteria (14.0%)
893,430 assigned reads
Virus (81.2%)
b
Other ssDNA virus (<0.1 %)
Unclassified virus (1.5%)
Circoviridae (13.8%)
Dicistroviridae (48.7%)
726,286 virus sequence reads
Parvoviridae (18.6%)
Other dsDNA virus (0.3%)
Podoviridae (2.7%)
Myoviridae (4.0%)
Siphoviridae (8.7%)
Nodaviridae (0.4%)
Picornaviridae (0.3%)
Leviviridae (0.2%)
Iflaviridae (0.1%)
Virgaviridae (<0.1%)
Caliciviridae (<0.1%)
dsRNA virus (<0.1%)
Other ssRNA virus (0.6%)
c
Plant host (<0.1%)
Others (2.3%)
Vertebrate host (14.1%)
Bacteria host (15.7%)
726,286 virus sequence reads
Invertebrate host (67.9%)


a
Partial nonstructural polyprotein
Structural polyprotein
ORF3
Calhevirus 2a
ORF1
ORF2
(A)n
GxxGxGKS
DD
H-D-S
KDELR
YGDD
FLKR
b
Aurantiochytrium single-stranded RNA virus 01 (NC_007522)
Chaetoceros socialis f.radians RNA virus 01 (NC_012212)
Chaetoceros tenuissimus RNA virus 01 (AB375474)
Rhizosolenia setigera RNA virus 01 (NC_018613)
Nora virus (DQ321720)
Infectious flacherie virus (NC_003781)
Deformed wing virus (NC_004830)
Sacbrood virus (NC_002066)
Aphid lethal paralysis virus (NC_004365)
Rhopalosiphum padi virus (NC_001874)
Black queen cell virus (NC_003784)
Himetobi P virus (NC_003782)
Triatoma virus (NC_003783)
Plautia stali intestine virus (NC_003779)
Homalodisca coagulata virus-1 (NC_008029)
Cricket paralysis virus (NC_003924)
Drosophila C virus (NC_001834)
Solenopsis invicta virus 1 (NC_006559)
Acute bee paralysis virus (NC_002548)
Israel acute paralysis virus (NC_009025)
Kashmir bee virus (NC_004807)
Solenopsis invicta virus 2 (NC009544)
Cherry leaf roll virus (NC_015414)
Tomato ringspot virus (NC_003840)
Cowpea mosaic virus (NC_003549)
Heterosigma akashiwo RNA virus (NC_005281)
Calhevirus 1 (HM480374)
Calhevirus 2a (AB937990)
Calhevirus 2b (AB937991)
Norwalk virus (NC_001959)
Vesicular exanthema of swine virus (NC_002551)
Picalivirus A (JQ898334)
Ljungan virus 87-012 (NC_003976)
Cadicivirus A (NC_021178)
Poliovirus (NC_002058)
Foot-and-mouth disease virus (NC_004004)
Hepatitis A virus (NC_001489)
0.2
Labyrnavirus
Bacillarnavirus
Iflaviridae
Dicistroviridae
Secoviridae
Marnaviridae
Caliciviridae
Picornaviridae


a
Partial 5'-UTR
P1
P2
P3
3'-UTR
3B
1
246
493
748
779
912
1047
1379
1476
1504
1699
2170
VP0
VP3
VP1
2A
2B
2C
3A
3C
3D
(A)n
KxKxxRxK
DxExNPGP
GxxGxGKS
DD
Y
H-D-C
GxCG
GxH
KDELR
GxPSG
YGDD
FLKR
b
Similarity between amino acid sequences (%)
100
80
60
40
20
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
Amino acid position
Crohivirus 1 (AB937989) vs
Human parechovirus 1 (L02971)
Ljungan virus 87-012 (NC_003976)
Sebokele virus 1 (NC_021482)
c
Human parechovirus 4 (AB433629)
Human parechovirus 7 (EU556224)
Human parechovirus 3 (GQ183028)
Human parechovirus 6 (EU077518)
Human parechovirus 1 (L02971)
Human parechovirus 2 (NC_001897)
Human parechovirus 5 (HQ696576)
Human parechovirus 8 (EU716175)
Ljungan virus 64-7855 (EU854568)
Ljungan virus M1146 (AF538689)
Ljungan virus 87-012 (NC_003976)
Ljungan virus 145SL (FJ384560)
Sebokele virus 1 (NC_021482)
Ferret parechovirus (KF006989)
Parechovirus
Seal picornavirus type 1 (NC_009891)
Aquamavirus
Turkey avisivirus (KC614703)
Avisivirus
Duck hepatitis A virus 1 (NC_008250)
Avihepatovirus
Crohivirus 1 (AB937989)
Swine pasivirus 1 (NC_018226)
Pasivirus PLV-CHN (JX491648)
Pasivirus
Hepatitis A virus (NC_001489)
Hepatovirus
Poliovirus (NC_002058)
Enterovirus
Porcine sapelovirus 1 (NC_003987)
Sapelovirus
Mosavirus (JF973687)
Mosavirus
Foot and mouth disease virus (NC_004004)
Aphthovirus
Cosavirus A (NC_012800)
Cosavirus
Seneca Valley virus (NC_011349)
Senecavirus
Miniopterus schreibersii picornavirus 1 (JQ814851)
Mischivirus
Encephalomyocarditis virus (NC_001479)
Cardiovirus
Equine rhinitis B virus 1 (NC_003983)
Erbovirus
Porcine teschovirus 1 (NC_003985)
Teschovirus
Bovine hungarovirus 1 (NC_018668)
Hunnivirus
Rosavirus (JF973686)
Rosavirus
Canine picodicistrovirus (NC_021178)
Dicipivirus
Turkey hepatitis virus (NC_021201)
Megrivirus
Turdivirus 2 (NC_014412)
Oscivirus
Turkey gallivirus (NC_018400)
Gallivirus
Turdivirus 1 (NC_014411)
Passerivirus
Aichi virus (NC_001918)
Kobuvirus
Salivirus (NC_012957)
Salivirus
0.2


a
Partial nonstructural polyprotein
Structural polyprotein
ORF3
Calhevirus 2a
ORF1
ORF2
(A)n
GxxGxGKS
DD
H-D-S
KDELR
YGDD
FLKR
b
Aurantiochytrium single-stranded RNA virus 01 (NC_007522)
Chaetoceros socialis f.radians RNA virus 01 (NC_012212)
Chaetoceros tenuissimus RNA virus 01 (AB375474)
Rhizosolenia setigera RNA virus 01 (NC_018613)
Nora virus (DQ321720)
Infectious flacherie virus (NC_003781)
Deformed wing virus (NC_004830)
Sacbrood virus (NC_002066)
Aphid lethal paralysis virus (NC_004365)
Rhopalosiphum padi virus (NC_001874)
Black queen cell virus (NC_003784)
Himetobi P virus (NC_003782)
Triatoma virus (NC_003783)
Plautia stali intestine virus (NC_003779)
Homalodisca coagulata virus-1 (NC_008029)
Cricket paralysis virus (NC_003924)
Drosophila C virus (NC_001834)
Solenopsis invicta virus 1 (NC_006559)
Acute bee paralysis virus (NC_002548)
Israel acute paralysis virus (NC_009025)
Kashmir bee virus (NC_004807)
Solenopsis invicta virus 2 (NC009544)
Cherry leaf roll virus (NC_015414)
Tomato ringspot virus (NC_003840)
Cowpea mosaic virus (NC_003549)
Heterosigma akashiwo RNA virus (NC_005281)
Calhevirus 1 (HM480374)
Calhevirus 2a (AB937990)
Calhevirus 2b (AB937991)
Norwalk virus (NC_001959)
Vesicular exanthema of swine virus (NC_002551)
Picalivirus A (JQ898334)
Ljungan virus 87-012 (NC_003976)
Cadicivirus A (NC_021178)
Poliovirus (NC_002058)
Foot-and-mouth disease virus (NC_004004)
Hepatitis A virus (NC_001489)
0.2
Labyrnavirus
Bacillarnavirus
Iflaviridae
Dicistroviridae
Secoviridae
Marnaviridae
Caliciviridae
Picornaviridae