



Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses

Raja Duraisamy¹ · Jean Akiana² · Bernard Davoust¹ · Oleg Mediannikov¹ · Caroline Michelle¹ · Catherine Robert¹ · Henri-Joseph Parra² · Didier Raoult¹ · Philippe Biagini³ · Christelle Desnues¹

Received: 11 August 2017 / Accepted: 16 February 2018 / Published online: 23 February 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Most of the emerging infectious diseases reported so far originated in wildlife. Therefore, virological surveillance of animals and particularly great apes is of great interest to establish the repertory of viruses associated with healthy hosts. This will further help to identify the emergence of new viruses and predict the possibility of interspecies transmission. In this study, we performed shotgun viral metagenomics on stool samples collected from seventeen free-living wild gorillas from the Republic of the Congo. The analysis revealed the presence of novel RNA viruses (picobirnaviruses, partitivirus, and *Picornavirales* (posa-like and dicistrovirus-like viruses)). Among these, picobirnavirus-related sequences were abundantly covered in the stools. Based on genetic variations both in capsid and RdRp proteins of picobirnaviruses, at least 96 variants were identified and most of them were novel. Among the 96, 22 variants had a nearly complete genome or segment. A comprehensive sequence analysis identified a potential new genogroup/genetic cluster and the presence of a short linear amino acid motif (ExxRxNxxxE) in a hypothetical protein. The sequence analysis of posa-like virus and dicistrovirus showed that these two viruses were novel members in the respective viral families. In conclusion, the identification of novel RNA viruses and their genetic diversity increases our knowledge about viruses that are associated with stools of wild gorillas and contributes to the initiatives in the search for potential emerging zoonotic viruses.

Keywords Wild gorillas · Viral metagenomics · Virus discovery · Novel picobirnaviruses · Novel picornavirus · Amino acid sequence motif

Edited by Detlev H. Kruger.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11262-018-1543-6>) contains supplementary material, which is available to authorized users.

✉ Raja Duraisamy
raja_king2k@yahoo.com

✉ Christelle Desnues
christelle.desnues@univ-amu.fr

¹ Aix-Marseille Université, CNRS 7278, IRD 198, INSERM 1095, Assistance-Publique des Hôpitaux de Marseille, Unité de Recherche sur les Maladies Infectieuses Tropicales Emergentes (URMITE), IHU Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France

² Laboratoire national de santé publique, Brazzaville, Republic of the Congo

³ Unité Mixte de Recherche 7268 ADES, Etablissement Français du Sang, Marseille, France

Introduction

The last decades have witnessed the advent and widespread application of viral metagenomics to study viral diversity and to describe novel viruses from many animal hosts, including non-human primates (NHPs) [1–4]. Such virological surveillance of viruses circulating in wildlife can serve as baseline exploratory dataset that can be further explored to identify the presence of unusual viruses in hosts showing symptoms of diseases. Such data are also critical to assess the changes in viral diversity in response to land-use and agricultural industry changes, as these are important environmental drivers of disease emergence. This could help us, in a close future, to take appropriate responses to mitigate those changes and potentially reduce zoonotic transmission [1]. Finally, virome comparison between different hosts will also be helpful to detect interspecies transmission [5].

Because humans and NHPs are genetically closely related, particularly those of the *Hominidae* family, the risk

of interspecies transmission of pathogenic microorganisms is higher. This is particularly critical for those who may have close contact with human populations as it may represent a threat to both humans and NHPs [5, 6]. For example, the human immunodeficiency virus (HIV) originated from chimpanzees and is responsible of a well-established infection in the human population. Another example of zoonotic transmission from great apes to humans is Ebola, with only sporadic events reported so far [7]. Conversely, human Metapneumovirus and Respiratory Syncytial virus are two examples of human to animal transmission, which have threat to great ape survival [8]. Special attention should thus be paid to detect and characterize novel viruses that may be of potential risk for interspecies transmission and subsequent infection. Previous virological surveys on NHPs have identified several human pathogenic viruses such as Ebola [5], Human Metapneumovirus [6], Respiratory Syncytial virus [8], and Picornavirus [9]. Gorillas, one of the four *Homini-dae* NHPs, are known to harbor many viruses [7]. Several studies had been conducted on gorillas from the African region, mostly targeting known pathogens by molecular and serological methods. These studies have identified many viruses such as Adenovirus (Gabon [10], Republic of the Congo [11]), Simian foamy virus (Cameroon, Gabon [12]), Herpesvirus (Republic of the Congo [11]), Human T-lymphotropic virus type 4 [13], Enteroviruses [14], Hepatitis B virus (Cameroon [15]), Human Metapneumovirus (Rwanda) [6], and Polyomavirus (Republic of the Congo) [16] from gorillas. Viral metagenomic studies on NHPs are scarce and have mostly targeted DNA viruses either on wild-living or captive animals. For instance, several novel circular ssDNA viruses were identified, while their pathological role remains to be determined [2, 17, 18]. Recently, a novel astrovirus strain, close to the human MLB2 strain, has been identified from stool of a diarrheic captive chimpanzee in China [3].

In order to expand our knowledge on the viral repertoire associated with free-living NHPs, stool samples were collected from wild gorillas in the Republic of the Congo, Central Africa in 2015. To our knowledge, this is the first report describing RNA viromes from wild gorilla feces. The stool samples were analyzed by shotgun viral metagenomics and sequences related to novel RNA viruses, i.e., novel picobirnaviruses, partitivirus, and *Picornavirales* (posa-like and dicistrovirus-like viruses) were described.

Materials and methods

Study site and gorilla population

The western lowland gorilla (*Gorilla gorilla gorilla*) is threatened with extinction due to habitat destruction, infectious diseases (especially the respiratory ones), and illegal

bushmeat trade. The Lésio-Louna and South-West Lésini Gorilla Reserves are part of a collaborative project between the Government of the Republic of the Congo and the Aspinall Foundation that manage a protected area of 170,000 ha that is located about 140 km north of Brazzaville (Fig. 1a, b). The aim of this project is to protect threatened ecosystems in Congo-Brazzaville, as well as species living there—particularly gorillas. About 35 gorillas are sheltered in the Lesio-Louna/Lésini Natural Reserves and in August 2015, a Franco-Congolese scientific mission was conducted to investigate and establish the repertory of the fecal microflora (bacteria, viruses, parasites) associated with this wild gorilla population. This mission was approved by the Ministry of Health (No 208/MSP/CAB.15 of 20 August 2015) and the Forest Economy and Sustainable Development (No 94/MEFDD/CAB/DGACFAP-DTS of 24 August 2015) of the Republic of the Congo. All the animals were apparently healthy.

Sample collection

For sample collection, gorillas (either isolated individuals or individuals in a group) were attracted by the reserve staff using apple fruits as baits. Fresh fecal samples (about 30 g) were then collected immediately after defecation from inside the stool with a sterile spatula, placed in a sterile collection tubes containing Saber medium and transported to the laboratory at Aix-Marseille University, France [19]. Fecal samples G1–G5 were collected from 5 isolated individuals. For the gorilla number 3 (G3), two samples were collected (G3A and G3B) on the same day. Samples G6–G17 (12 samples) were retrieved from different stools collected from a group of gorillas (about 30 individuals) and could not be associated with any particular gorilla. Table 1 provides information about the samples and the individuals. The export of the samples to France was authorized by the Ministry of Health (No 136/MSP/DGELM/DHP/SHE of 28 August 2015). In the laboratory, the samples were stored in aliquots at -80°C , until they were processed.

Sample preparation

The eighteen fecal samples (1 g) were diluted (10–20%) in HBSS (Hanks Basic Salt Solution), vortexed, centrifuged and the supernatants were filtered through 0.45- μm filters (Millipore). To remove unprotected nucleic acids, the filtrates were first treated with 1 μl of RNase A (Ambion) for 15 min, followed by 1 μl of Benzonase (Novagen) and 4 μl of Turbo DNAase (Ambion) for 45 min, as described elsewhere [20]. The samples were extracted in two batches. First, the total nucleic acids (TNAs) were extracted from all 18 samples with QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA) and the extracted nucleic acids were pooled (pool 1). In

Fig. 1 a Worldwide detection of picobirnavirus sequences. World map highlighting the 23 countries (red bullets) from which picobirnavirus genetic sequences have been reported based on both published articles and sequences available in GenBank. Each country is named with its three-digit International country code. **b** Map and description of the sampling site. Country map of the Republic of the Congo, highlighted with the appropriate place of sample collection. The original templates were downloaded from the public domain. (<https://commons.wikimedia.org/wiki/File:BlankMap-World-alt.png#filelinks>; https://commons.wikimedia.org/wiki/Atlas_of_the_Republic_of_the_Congo#/media/File:Congo_republic_sm04.png) (Color figure online)

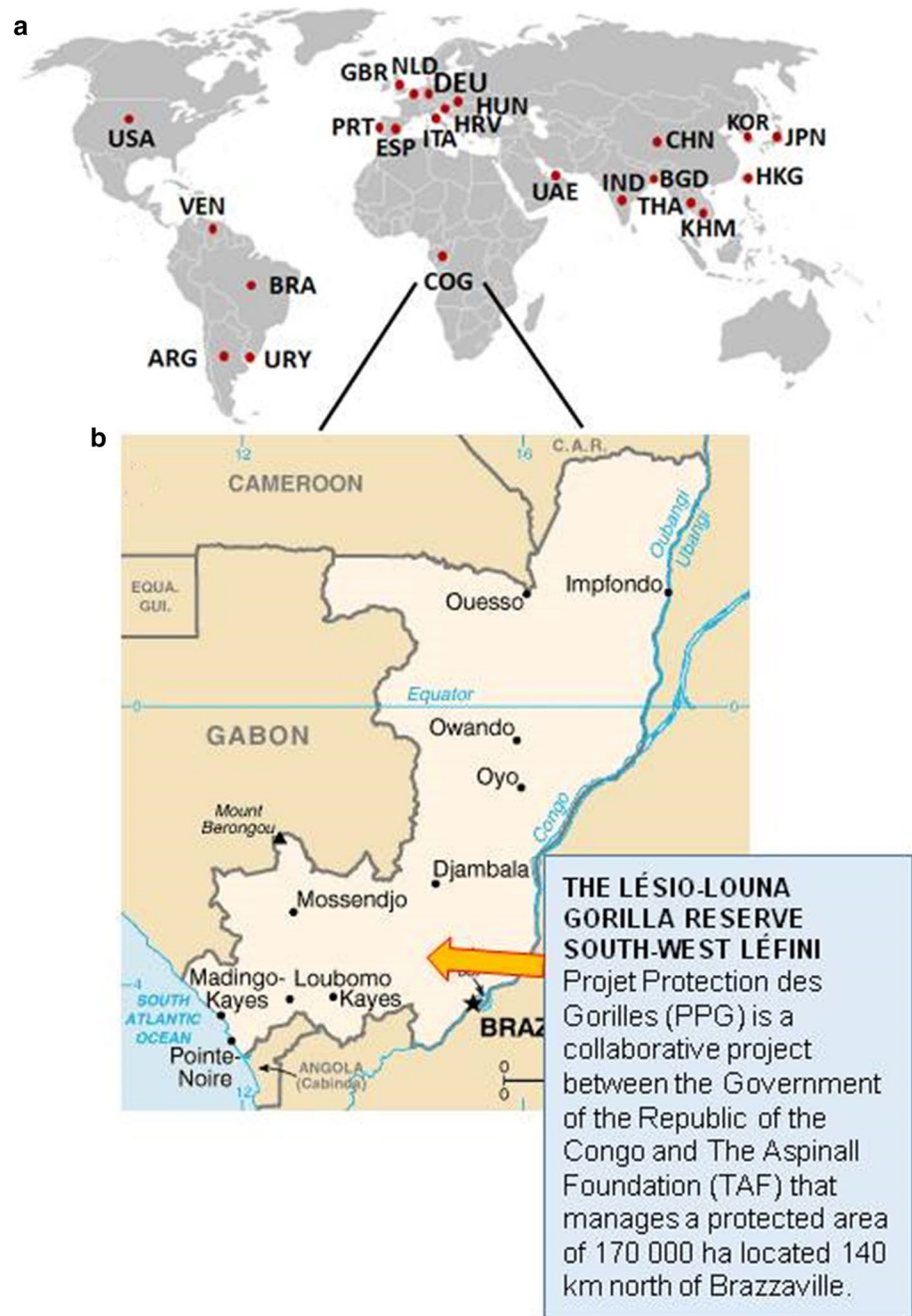


Table 1 Description of the gorillas (name, sex, age when applicable) and samples (sample Id., nature, coordinates, and sampling date)

Gorilla name	Description	Sample id	Age (years)	Sex	Sample nature	Coordinates	Sampling date
Yengo	Isolated individual	G1	3	Male	Fresh	3°16'16"S–15°28'12"E	2015-08-26
Owando	Isolated individual	G2	4	Male	Fresh	3° 16'16"S–15° 28'12"E	2015-08-26
Kelle	Isolated individual	G3A, G3B	17	Male	Fresh	3° 07'53"S–15° 31'30"E	2015-08-27
Sid	Isolated individual	G4	35	Male	Fresh	3° 02'23"S–15° 30'23"E	2015-08-28
John	Isolated individual	G5	30	Male dominant	Fresh	3° 01'56"S–15° 28'34"E	2015-08-28
–	About 30 individuals (John's group)	G6 to G17	–	n/a	Fresh	3° 01'56"S–15° 28'34"E	2015-08-28

a second batch, all the samples were extracted again with the Roche High Pure Viral Nucleic Acids Extraction Kit (Roche) and formed another five pools: pool 2 (G3A and G3B), pool 3 (G1–G3B), pool 4 (G4–G7), pool 5 (G8–G12), and pool 6 (G13–G17) to obtain more sequences from these samples.

RNA viral library preparation for Illumina sequencing

The TNAs were reverse-transcribed using SuperScript III Reverse Transcriptase (Invitrogen) and Random Hexamers (Invitrogen). The samples were barcoded in order to be mixed with 15 other metagenomic projects constructed by the Nextera XT DNA sample prep kit (Illumina). The libraries were prepared according to the Nextera XT protocol (Illumina). The libraries were purified, normalized, and then pooled (six pools) for sequencing on a MiSeq Platform (Illumina Inc, San Diego, CA, USA) using paired-end strategy. Automated cluster generation and paired-end sequencing with dual index reads were performed in a single 39-h run in 2×250 bp.

Sequence analysis

The Illumina sequences were analyzed using CLC genomics V7.5 (www.clcbio.com). The paired-end Illumina reads were imported into CLC and the adaptor was removed. The low-quality sequences were trimmed and then the cleaned sequences were de novo assembled using default parameters. The resulting contigs were subjected to BLASTx against nr database (NCBI) with a selection of “virus” as organism (the feature available in the CLC to blast the sequences against a virus database). Contigs with similarities to viral sequences of picobirnaviruses, posa-like viruses, dicistroviruses, and a partitivirus were detected. The related contigs were retrieved from all the sequencing pool for reconfirmation; the retrieved sequences were subjected again to BLASTx against nr database (NCBI) and the deduced protein sequences were re-confirmed with protein search (BLASTp) against nr database and appropriate classification.

Phylogenetic analysis

The putative Open Reading Frames (ORFs) were predicted by NCBI ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder/>). The deduced viral protein sequences were aligned along with representative sequences of respective viruses which were retrieved from GenBank using MEGA V.5.1 [21] or V6.1 [22]. Then, the phylogenetic trees were constructed by maximum likelihood methods with best substitution model selected using ProtTest and 100 bootstrapped iterations.

Pairwise identity was calculated using sequence demarcation tool (v1.2) [23] and MEGA v6.1 [22]. The logo of ExxRxNxxxE amino acid motif was generated with <http://weblogo.berkeley.edu> with default parameter.

Recombination analysis

Complete RdRp coding sequences were examined for recombination using Simplot [24] and Rat [25], according to the default parameters.

Nucleotide sequences: All the sequences generated from this study were deposited in GenBank under the accession number KY502835-KY503023.

Result

Sequencing result

Total information of 7.4 Gb was obtained from a 784 K/mm² cluster density with a cluster passing quality control filters of 94.5% (1,509,700 clusters). Within this run, the index representations of the 6 pools were determined to be of 1.56, 1.54, 7.32, 3.01, 5.89, and 5.39%, respectively. The raw data of 222,210; 220,126; 1,044,848; 428,865; 840,562; and 769,080 paired-end reads (total of 3,525,691 paired-end reads) were trimmed and filtered according to the read qualities.

Picobirnaviruses

After annotation by BLASTx, only contigs with similarities to RNA viruses were analyzed. From the six sequencing pools, 339 contigs related to picobirnaviruses were retrieved for further analysis. Finally, 289 contigs were re-confirmed by BLASTx against the nr database (NCBI) as picobirnaviruses. The sequences were grouped into 96 variants based on their genetic (either capsid [Segment 1] or RNA-dependent RNA polymerase (RdRp) [Segment 2]) relationship by pairwise identity analysis, and the maximum percentage of identity observed between the strains was 70%. These variants were present in more than one sequencing pool. The details of variants and its genetic relationship are presented in Table 2 and Supplemental Table 1. The variants are tentatively named as “Institut Hospitalo Universitaire-Congo-Gorilla-picobirnavirus Variants and abbreviated as “IHU-Con-GPbvs-V1 to 96.” Henceforth, these viruses are referred to as variants (GPbvs V1–V96) when appropriate in the text.

Genetic compositions

Among the 96 variants, 22 (15 based on Seg 1 and 7 based on Seg 2) have nearly complete segments with 13

Table 2 Description of nearly complete genome segments of gorilla picobirnaviruses

S. no	Segment	Variants name	Contig		No of base pairs	ORF	No of amino acids	Protein weight (Kd)
			Number	Total read count/average coverage (X)				
1	Seg1	IHU-Con-GPbvs-V1	ContigS1-2349	194/15	2916	Orf 1-220-950 bp Orf 2-982-2697 bp	244 571	63.87
2	Seg1	2 IHU-Con-GPbvs-V2	ContigS4-29	25541/1998	2793	Orf 1-384-908 bp Orf 2-905-2581 bp	174 558	62.72
3	Seg1	IHU-Con-vs-V3	ContigS4-1249	80/6	2712	Orf 1-125-820 bp Orf 2-880-2559 bp	231 559	62
4	Seg1	IHU-Con-GPbvs-V4	ContigS1-400	426/36	2671	Orf 1-165-1046 bp Orf 2-1059-2627 bp	293 522	57.89
5	Seg1	IHU-Con-GPbvs-V5	ContigS4-111	1596/134	2646	Orf 1-80-934 bp Orf 2-945-2573 bp	284 542	59.76
6	Seg1	IHU-Con-GPbvs-V6	ContigS5-821	881/75	2587	Orf-1-487-810 Orf-2-830-2419	107 529	58.73
7	Seg1	IHU-Con-GPbvs-V7	ContigS3-730	1588/139	2583	Orf 1-165-710 bp Orf 2-728-2524 bp	181 598	66.78
8	Seg1	IHU-Con-GPbvs-V8	ContigS4-1022	254/22	2572	Orf 1-389-967 bp Orf 2-981-2552 bp	192 523	59.58
9	Seg1	IHU-Con-GPbvs-V9	ContigS1-1263	60/5	2476	Orf 1-312-722 bp Orf 2-728-2431 bp	136 567	62.77
10	Seg1	IHU-Con-GPbvs-V10	ContigS1-446	406/37	2468	Orf 1-398-736 Orf 2-753-2408 bp	112 551	61.24
11	Seg1	IHU-Con-GPbvs-V11	ContigS6-563	147/13	2448	Orf 1-315-638 bp Of 2-631-2355 bp	107 574	64.78
12	Seg1	IHU-Con-GPbvs-V12	ContigS6-298	2232/208	2400	Orf 1-50-742 bp Orf 2-733-2340 bp	230 535	60.09
13	Seg1	IHU-Con-GPbvs-V13	ContigS5-66	3354/316	2349	Orf 1-103-444 bp Orf 2-499-2103 bp	113 534	58.51

Table 2 (continued)

S. no	Segment	Variants name	Contig		No of base pairs	ORF	No of amino acids	Protein weight (Kd)
			Number	Total read count/average coverage (X)				
14	Seg1	IHU-GPbvs-V14	ContigS5-17036	135/13	2344	Orf 1-22-687 bp Orf 2-689-2302 bp	221 538	59.53
15	Seg1	IHU-Con-GPbvs-V15	ContigS3-2734	1887/189	2208	Orf 1-203-394 bp 405-2177 bp	66 590	64.82
16	Seg 2	IHU-Con-GPbvs-V16	ContigS1-111	181/17	2355	Orf 3-685-2304	539	61.45
17	Seg 2	IHU-Con-GPbvs-V17	ContigS3-401	165/16	2298	Orf 3-626-2257	543	62.26
18	Seg 2	IHU-Con-GPbvs-V18	ContigS5-1203	3829/388	2204	Orf2-555-2114	519	59.53
19	Seg 2	IHU-Con-GPbvs-V19	ContigS1-1231	271/30	2045	Orf 3-408-1997	529	60.51
20	Seg 2	IHU-Con-GPbvs-V20	ContigS6-233	107/12	1991	Orf-195-1862	555	63.88
21	Seg 2	IHU-Con-GPbvs-V21	ContigS1-141	309/41	1718	Orf061-1659	532	60.9
22	Seg 2	IHU-Con-GPbvs-V22	ContigS3-872	1689/221	1715	Orf-57-1694	545	63.77

S. no	Genetic relationship of GPbvs with other strains %					Frequency of detection in sequencing pools		
	Query coverage	Identity (e-value)	Similarity	Most similar virus	Genbank accession no.	No of sequences	Contig number	Genbank accession number
1	95	27(2e-43)	44	Dromedary (AIY31265)	KY502835	2	S5-151 (93%)	KY502931
3	99	33 (2e-76)	49	Dromedary (AIY31275)	KY502836	4	S1-2308 (99 %), S6-1719 (99%), S5-2580 (99%)	KY502932- 34
4	72	32 (1e-48)	50	Dromedary (AIY31272)	KY502838	2	S5-111 (100%)	KY502935
5	95	32 (2e-74)	50	Horse (AKN50623)	KY502839	1		
6	59	31 (1e-37)	51	Dromedary (AIY31279)	KY502840	5	S6-499 (99%), S1-132 (78%), S4-34361 (96%); S1-17168 (100%)	KY502936-39
7	99	36(2e-109)	51	Porcine (YP_009241385)	KY502841		S5-50720, S5-17943, S1-820-99%	KY502940-42
8	92	26 (1e-35)	42	Dromedary (AIY31281)	KY502842	3	S1-319 (99%), S5-1184 (100%)	

Table 2 (continued)

S. no	Genetic relationship of GPbvs with other strains %					Frequency of detection in sequencing pools		
	Query coverage	Identity (e-value)	Similarity	Most similar virus	Genbank accession no.	No of sequences	Contig number	Genbank accession number
9	56	29 (1e–28)	46	Human (YP_239360)	KY502843	1		
10	81	39 (1e–95)	57	Human (AIG71989)	KY502844	3	S5-591 (99%), S6-406 (99%),	KY502945-46
11	98	31 (3e–76)	49	Human (A14000957)	KY502845	5	S1-1103 (100%), S5-4718 (100%), S1-6773 (99%), S5-35532 (99%)	KY502947-50
12	96	30 (9e–66)	48	Dromedary (AIY31265)	KY502846	3	S1-712 (100%), S5-80 (99%)	KY502951-52
13	70	33 (2e–46)	49	Dromedary (AIY31282)	KY502847	2	S1-1575 (99%),	KY502953
14	99	39 (3e–119)	53	Otarine (AMP1 8946)	KY502848	7	S1-1330 (100%), S6-49225 (99%), S6-20144 (100%), S4-6355 (98%), S6-29669 (99%), S1-16898 (100%)	
15	84	29 (4e–51)	49	Dromedary (AIY31277)	KY502849	2	S1-6400 (99%)	KY502960
16	95	47(6e–170)	66	Wolf (ANS53885)	KY502850	2	S4-400 (99%),	KY502961
17	99	60 (0)	75	Dromedary (AIY31288)	KY502851	1		
18	72	52(1e–173)	67	Chctun (A1dV53314)	Kf 502852	3	S4-188 (99%), S1-1434 (100%)	
19	99	67(0)	80	Human (AAG53583)	KY502853	4	S3-179 (99%), S4-2826 (99%), S5-3794 (99%)	KY502964-66
20	96	25(1e–19)	38	Horse 4 (AKN50618)	KY502854	2	S5-398 (99%)	KY502967
21	92	73 (0)	84	Wolf (ANS53885)	KY502855	3	S5-986 (99%), S6-524 (99%)	
22	89	48(5e–163)	74	Monkey (AQS16638)	KY502856	2	S1-651 (99%)	KY502970

Contig name: it starts with sequencing pool number-then contig number

IHU-Con-GPbvs V Institut Hospitalo Universitaire-Congo-Gorilla-picobirnavirus Variants, *Seg* segment, *ORF* open reading frame, *ORF 1* hypothetical protein, *ORF 2* capsid, *ORF 3* RdRp (RNA-dependent RNA polymerase)

of them having a sequencing coverage > 30X. Each segment was covered with numerous reads with a range from 60 to 25,541 (Table 2 and Supplemental Fig. 1). Of the 22 variants, 18 (GPbvs V1, V3, V4, V6–V8, V10–V16, V18–V22) were found in more than one sequencing pool (Table 2). The Supplemental Fig. 1 shows the coverage of each contig after re-mapping of the reads against its respective original contigs. The GC content of both proteins was analyzed along with strains detected from different hosts. The number of host and sequences analyzed

for the capsid are gorilla (16), human (3), rabbit (1), otarine (1), turkey (4), porcine (5), horse (4), dromedary (6), fox (1), and for the RdRp are gorilla (14), human (12), bovine (1), monkey (1), otarine (1), porcine (3), turkey (1), dromedary (4), and horse (4). The capsid proteins of GPbvs ($n = 16$) have an average GC content of 40%, and for the other strains, the GC content varies between 40 and 50% (Supplemental Fig. 2a). In case of the RdRp, the GPbvs have an average GC content of 43.6% and

similar values were observed from PBVs detected from other hosts (Supplemental Fig. 2b).

Segment 1 (long segment)

Fifteen variants had a nearly complete segment. The size of the segments ranged from 2208 to 2916 bp, which encode two open reading frames (ORF) and non-coding region at both terminals. Some variants (V-2, V-11, and V-12) had overlapping reading frames but most of them harbored an intergenic region between ORFs. In these intergenic regions, 5 to a maximum of 59 nucleotides were observed among the variants analyzed (Table 2 and Fig. 2a). The non-coding region at the 5' end had a maximum of 486 nucleotides and 246 nucleotides at the 3' end.

The ORF 1 encodes a protein (hypothetical) with an unknown function. The size of the proteins varies among the strains. The smallest protein (66 amino acids) was observed in V-15 and V-4 had the largest protein (293 amino acids) (Fig. 2a). The proteins are highly disordered in nature; a range of 15–68% was observed between the strains (Fig. 2b). The strains detected from humans and rabbits (recognized type species by the ICTV) had 40 and 54% of disordered protein, respectively. The homology search in the protein data bank (PDB) could not find any significantly related one. The hypothetical protein has short (10 amino acids) repetitive motifs (ExxRxNxxxE), which are observed in all PBVs detected from different hosts (Fig. 3a, b). The number of motifs varied between the variants; V-1 (S1-2349) had only 2 motifs, whereas up to 11 were observed in V-5 (S4-111) (Fig. 3b). The frequency of occurrence of the amino acids in the motifs is presented as a logo or graphical format in Fig. 3c. In the “X” position, three different amino acids can occur. The frequency of these three amino acids was calculated independently and included in the logo (highlighted within the box with the most frequent one placed at the first place). The final motif is E[TNS][KNA]R[HSA]N[LVR][AEQ][TKQ]E.

The ORF 2 encodes the capsid protein which size ranged from 522 to 598 amino acids in length. The preliminary analysis revealed that all these variants were novel and shared < 40% amino acid identity with other known PBV sequences available in GenBank (Table 2). The most divergent virus is V-1, which shared only 27% of identity with dromedary picobirnavirus (AIY31265). The capsid protein was highly divergent; particularly, the first 88 amino acids at the N-terminal region were hyper variable and quite distinct between each other's and had no homolog in the public database (Fig. 2b). Also, this defined region was almost completely disordered in nature.

Segment 2 (small segment)

Seven variants have a nearly complete segment. As observed with other known PBVs, the segment 2 displays a single ORF, which encodes the RdRp and harbors non-coding regions at both ends. The segment size ranges from 1.7 to 2.3 kb and encodes proteins ranging from 519 to 555 aa in length (Table 2; Fig. 2c). The variants share a range of 25 to 68% of identity with other known PBVs in GenBank. V-20 shares only 25% of identity with a strain detected from a horse (AKN50618). The Seg 2 has a long non-coding region ranging from 684 nucleotides (684 nt) at the 5' end of V-16 to 129 nt at the 3' end in V-20 (Table 2).

Complete RdRp coding sequences were analyzed with that from other PBVs available in GenBank. The RdRp proteins had two conserved amino acid motifs (GDD and SGxxxT). The pairwise identity analysis revealed that gorilla picobirnaviruses are distinct between each other and share a range of 20–70% of identity (Supplemental Fig. 3). Also, the mean pairwise identity of GPbvs with other picobirnaviruses ranged from 36 to 52%; strains detected from horse are quite distinct from GPbvs as well as with all other PBVs (Supplemental Table 2). A maximum likelihood phylogenetic tree built with all the strains revealed two genetic clusters (earlier referred to as genogroups) [26, 27]. The strain V-20 (S6-233) formed a separate branch (Fig. 4), sharing only 25% of identity with other strains in the public database (Table 2). This suggests further the existence of more genogroups or genetic clusters of PBVs. The genetic distance plot of nearly complete RdRp nucleotide sequences suggests that the strains detected from gorillas maintain an almost similar percentage of identity in the entire genomic positions which were analyzed (Fig. 5).

Relationship of gorilla strains closely related to human viruses

Two variants detected from gorilla GPbvs V-31 (1918 nt-RdRp) and GPbvs V-39 (1638 nt-RdRp) shared a high level of amino acid sequence identity with strains detected from humans during 2007–2008 in India. The V-31 shared 98% (BAJ53295) and V-39, 92% (BAJ53293) with these strains. Another three RdRp sequences, V-68 (494 nt), V-77 (383 nt), V-96 (224 nt), shared over 90% of identity (amino acids) with strains detected from *Macaca mulatta* in Bangladesh (Supplemental Table 1).

As previous studies demonstrated recombination events in PBVs [1], the gorilla sequences were analyzed along with sequences obtained from humans, monkeys, porcine, dromedaries, turkeys, and bovines. The analysis indicated that there was no apparent recombination event observed with the strains from gorilla.

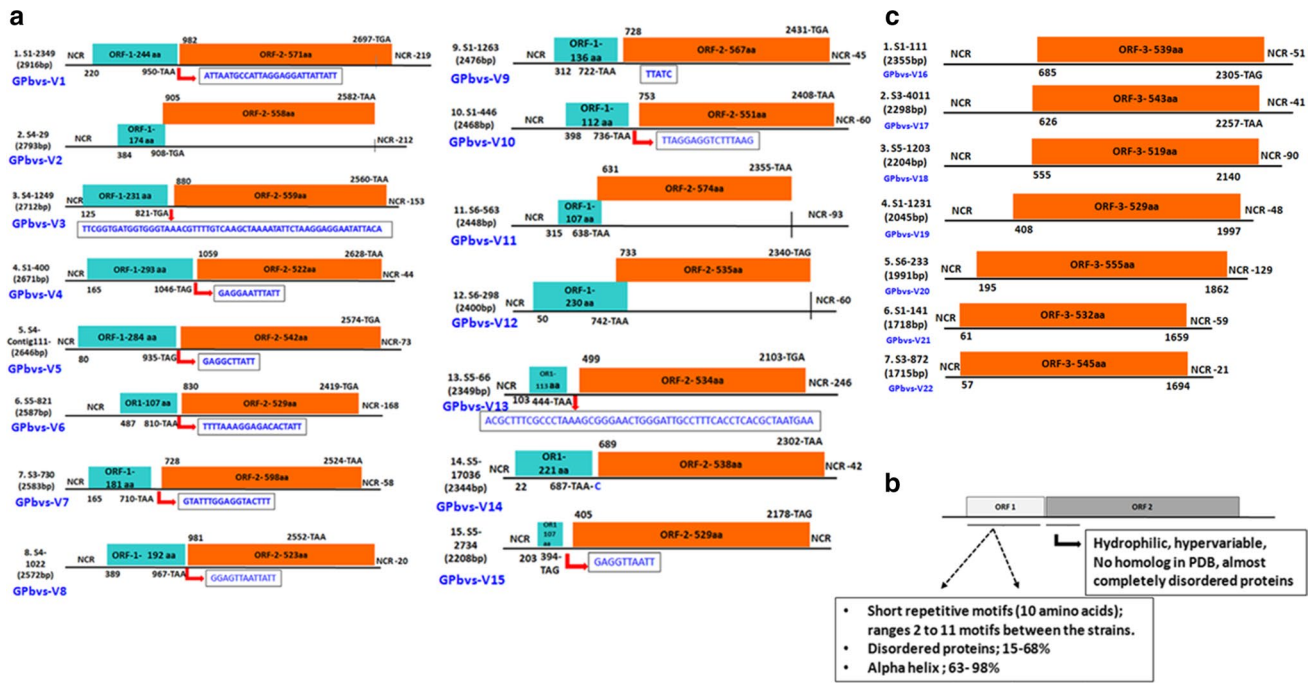


Fig. 2 Schematic representation showing the genomic organization of complete segments of GPbvs. **a** Genomic organization of the complete Segment 1 of 15 variants detected in gorilla stools. All the variants predicted with two ORFs (indicated by blue and orange boxes) and non-coding regions at both terminals. Three variants (V2, V11, and V12) have overlapping coding regions. The other 12 variants have an intergenic region between the ORFs. The nucleotide sequences in the intergenic region are presented in the box. Each variant is labeled with its original contig name and number of bp within brackets. NCR—Non-coding region. Each ORF is highlighted with

starting and ending coding region, number of amino acids, and stop codon at the end. **b** Schematic representation of the genomic features observed in Segment 1 of all the 15 variants. The features observed from PBVs detected from different hosts are highlighted in the boxes. **c** Genomic organization of the complete Segment 2 of 7 variants obtained from gorilla stools. Genomic organization of Segment 2, which encodes one ORF and has non-coding regions at both terminal ends. The name of each variant and the annotation is the same as the one described in (a) (Color figure online)

Other novel viruses

Novel gorilla stool-associated partitivirus

Partitiviruses are non-enveloped, bi-segmented dsRNA viruses (1.4–2.4 kb in size). The smaller segment usually codes for the capsid protein and the larger usually codes for the virion-associated RNA polymerase. The viruses are associated with latent infections of their fungal, protozoan, and plant hosts [28]. There are no known natural vectors. Currently, the family *Partitiviridae* consists of five genera (*Alphapartitivirus*, *Betapartitivirus*, *Cryspovirus*, *Deltapartitivirus*, *Gammapartitivirus*) and 15 unassigned species (<http://www.ictvonline.org/virusTaxonomy.asp>).

In this study, a contig measuring about 2421 bp was identified (GenBank accession KY503021) in single sequencing pool (pool 5). The segment has one ORF, which encodes a putative RdRp protein and a non-coding region at both ends (210 bp at 5' end and 279 bp at 3' end). A BLASTx search against the Non-redundant protein database (NCBI) returns hit to Beet Cryptic Virus 3 (e-value $4e-05$) with only 25% of identity (query coverage of 27%). In order to

describe it effectively, the sequence was tentatively named as 'Gorilla Stool-associated Partitivirus' and abbreviated as GOSP. The GOSP sequence was analyzed along with representative sequences of other partitiviruses, available in GenBank. The alignment of amino acid (401 aa) sequences of 20 strains revealed that there are two motifs (KxR, GxPSG) and few conserved amino acids to this viral family. A maximum likelihood phylogenetic tree indicated that this strain was distinct and diverged from other strains with over 80% at the amino acid level and formed a separate branch (Fig. 4).

Novel gorilla stool-associated picornavirus

According to the ICTV, the order '*Picornavirales*' consists of five families (*Dicistroviridae*, *Flaviridae*, *Marnaviridae*, *Picornaviridae*, *Secoviridae*, and 2 unassigned genera). In this study, a contig about 2 kb (2281 bp) was identified (KY503022) in three sequencing pools (pools 4, 5, and 6). A conserved domain search of translated protein sequence gave hit to 'rhv_like picornavirus capsid protein' domain. The sequence detected from gorillas was then tentatively named



Fig. 3 Prediction of conserved, repetitive amino acid motifs in the ORF 1 of all picobirnaviruses. **a** The details of ORF 1 (hypothetical protein) of 26 variants (15 GPbvs and 11 PBVs from other hosts) were analyzed. This hypothetical protein shows repetitive amino acid motifs (ExxRxNxxxE) in all the variants. The motifs are highlighted with individual boxes. **b** Detailed version of (a) which clearly presents the individual motifs along with arrangement pattern. **c** Graphical

representation of the amino acid motifs (ExxRxNxxxE) and rep-
 aration of its ‘x’ amino acids. The bit size of each amino acid was
 directly related to the frequency of its presence (bit 4 is 100%). For
 the remaining amino acids (x), three amino acids are detected at each
 place and the most frequent among the three is presented at the top,
 as highlighted in the box

as GOSA (Gorilla Stool Associated) virus. The search of translated nucleotide sequences against the protein database showed that this virus was closely related to ‘posa-like viruses’ [29], *Husaviruses* [30], *Fisavirus* [31], *Rasavirus*, and *Basavirus* [32] and shared < 37% of amino acid identity with them. These viruses have been recently detected from porcine, humans, fishes, rats, and bats, respectively, and proposed as a new viral family in the order of *Picornavirales*, yet to be recognized by the ICTV. A phylogenetic tree was constructed based on 314 amino acids along with the most closely related viruses and representative sequences of all the members in the *Picornavirales*. The ‘posa-like viruses’ formed a separate cluster, and within the cluster, many groups were observed. The GOSA virus formed a separate branch and was distinct from other strains; the most closely related viruses are ‘posavirus 1-like viruses’ (Fig. 6a).

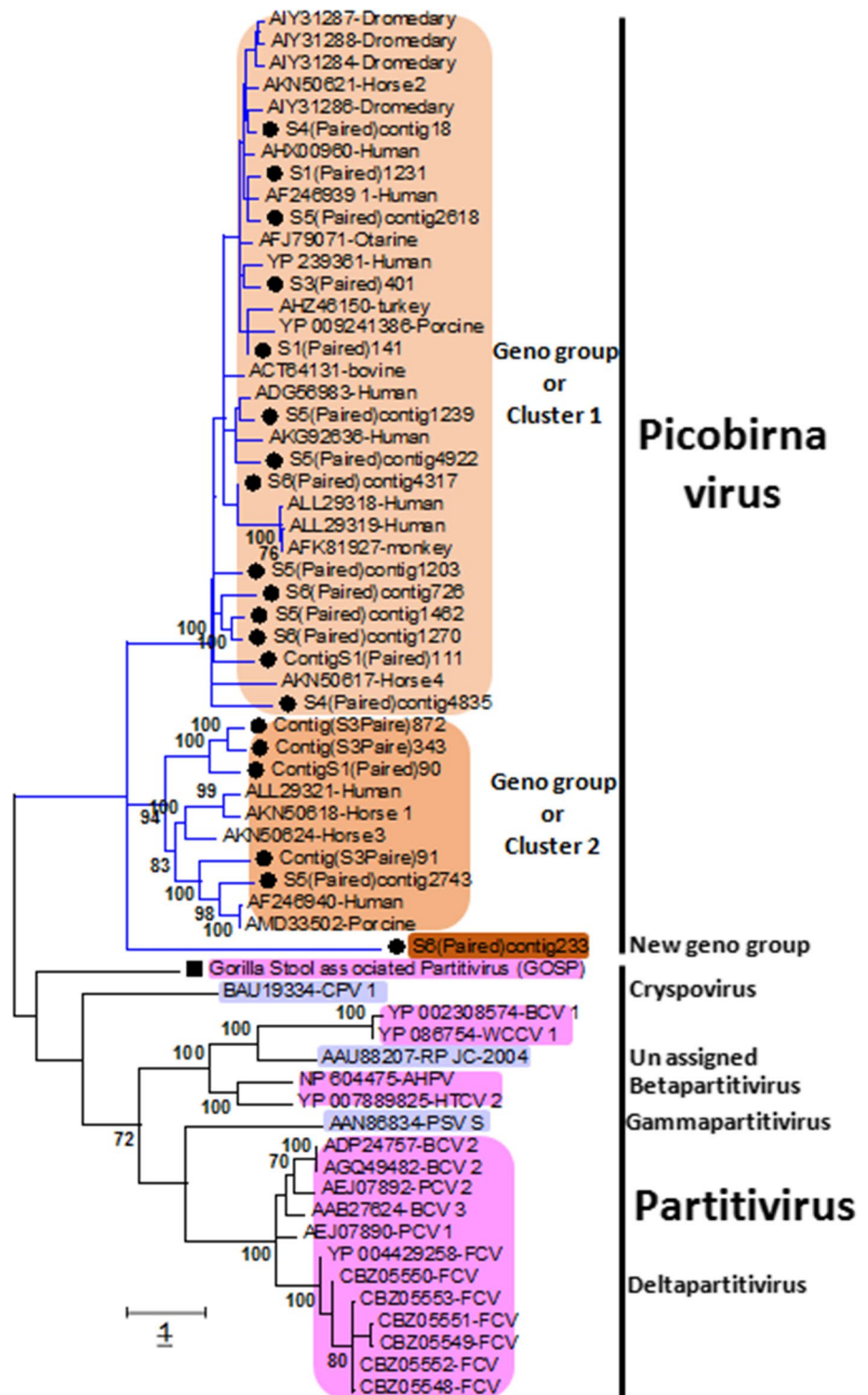
Novel gorilla stool-associated dicistrovirus

Dicistroviruses are another family in the order of *Picornavirales*, with a genome of about 8.5–10 kb in size. Currently, the family *Dicistroviridae* consists of 15 species, placed under three genera (*Aparavirus*, *Cripavirus*, *Triatovirus*). The genome encodes two open reading frames; the ORF 1 encodes an helicase, a viral protein, a protease, and a RNA-dependent RNA polymerase (pol), while ORF 2 encodes a capsid protein [33].

From gorilla stool samples, a contig of about 4,045 nucleotides was identified from single sequencing pool (pool 1), which has partial sequences of both ORFs (KY503023). The initial BLASTx search against NCBI nr database suggested that the ORF 1 (541 amino acids) was related to Northwest Territories Cripavirus with 48% identity (63% query coverage), followed by 33% of identity (95% coverage) with drosophila C virus. The capsid protein (710 amino acids) was highly divergent and merely shared 29% of identity with *Aphid lethal paralysis virus*. A nucleotide (BLASTn) search against nr/nt database of NCBI turned out to be a non-viral hit.

Both proteins (capsid 600 amino acids and RdRp 508 amino acids) were further analyzed along with other strains in the family. The alignment of RdRp (1245 to 1765 aa position of *Cricket paralysis virus* [NP_647481]) sequences identified many conserved amino acids and motifs (AGDxxxxD; PSGxxxTxxxN; GDD) common to the *Dicistroviridae*. The motif ‘GDD’ is also observed in the RdRp protein of picobirnaviruses and partitiviruses (except *Radish partitivirus* AAU88207). The pairwise distance analysis indicated that these strains were distinct from others with a range of 50–72% (mean distance of 62%). In the capsid protein, some conserved amino acids and motifs (PxxS, KTxHxxR) were observed. The pairwise distance analysis suggested that the gorilla stool-associated dicistrovirus was distinct from others by a minimum of 70% (*Aphid lethal*

Fig. 4 Phylogenetic analysis of picobirnaviruses and partitivirus detected from gorilla stools and other hosts. A maximum likelihood phylogenetic tree was constructed based on the RdRp proteins of the viruses with an algorithm with the lowest BIC score. The number of strains was gorilla-20, human-9, horse-4, dromedary-4, porcine-2, 1 each for turkey, otarine, bovine, and monkey. The variant names are labeled first with the GenBank accession number, followed by the host name. In the case of gorilla picobirnaviruses, the contig numbers were used as presented in Table 2 and Supplemental Table 1. The genetic clusters and branches are highlighted with different colors (Color figure online)



paralysis virus) to a maximum of 78% (*Taura Syndrome Virus*) with a mean distance of 70%. A maximum likelihood phylogenetic tree was constructed for both proteins; the gorilla stool-associated dicistrovirus formed a separate branch between the genus of *Aparavirus* and *Cripavirus* in both trees. The RdRp sequence aligned with Northwest

Territories Cripavirus (Fig. 6b, c). In all, although this gorilla stool-associated dicistrovirus was related to the genus of *Cripavirus*, it was quite distinct and could represent as novel member of this family.

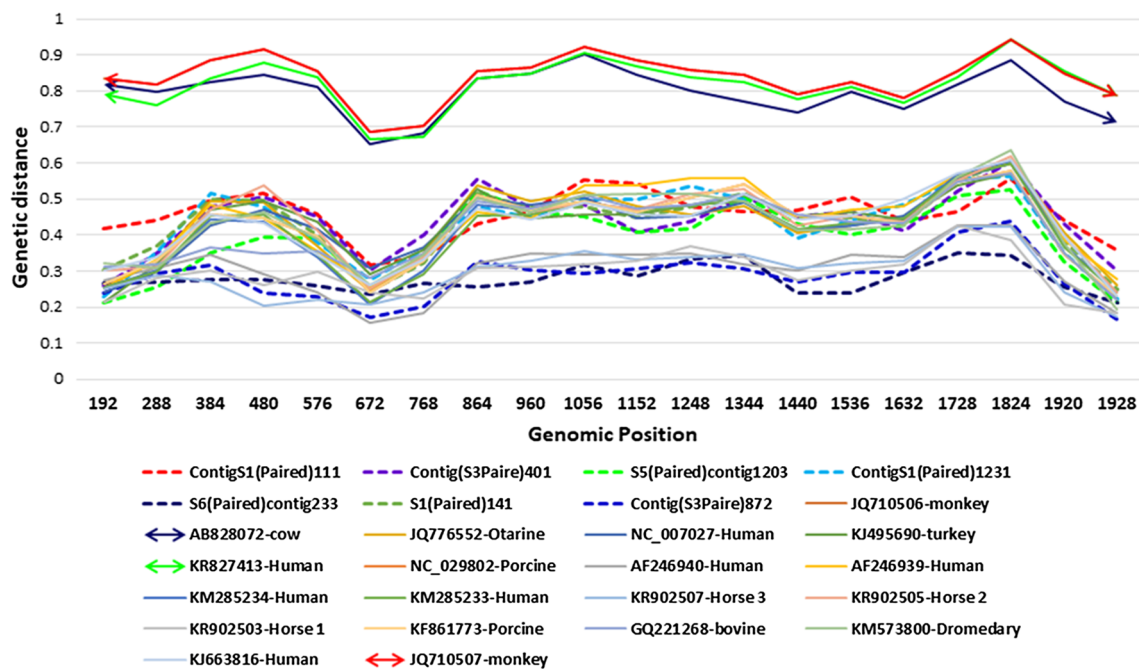


Fig. 5 Divergence plot of the RdRp sequences of picobirnaviruses. In this analysis, 26 picobirnaviruses detected from gorilla, human, porcine, monkey, cow, otarine, dromedary, turkey, and horse were

included. The seven GPbvs are highlighted with dotted lines. The “X” axis indicates the genomic position of the viral genome and the “Y” axis indicates the percentage of diversity

Discussion

Picobirnaviruses (PBVs) belong to the genus *Picobirnavirus* in the family *Picobirnaviridae*. These viruses were first identified from human [34] and rat stools in 1988 from Brazil [35]. Since then, these viruses have been detected from various symptomatic and asymptomatic hosts, i.e., human [36], domestic animals (bovine/cows/calf [37], cat [38], chicken, dogs [39], mouse [40], dromedary [41], pigs [42], rats [39], rabbit [43], horse/foals [44], turkey [45], wild animals (fox/red fox) [46], jaguar, lion [47], monkey [48], wild rodent [40], orangutan [49], puma [47], rhesus macaques [1], seal lion [50], small carnivores [51], snakes [39], wolf, bats (mammals) [52], and also in waste water [53]. However, only partial sequences are available for many of these viruses. To our knowledge, this is the first report that describes novel PBVs in wild gorilla stools (Fig. 1a, Supplemental Table 3).

Picobirnavirus particles are non-enveloped, about 33–37 nm in diameter enclosing a bi-segmented RNA genome [43]. The segment 1 is about 2.4–2.6 kb and segment 2 ranges from 1.5 to 1.9 kb. The analysis of nearly complete sequences (GPbvs) of both segments indicates that the nucleotide composition and GC contents varies between strains detected from different hosts (Supplemental Fig. 2a, b). For each protein, strains from nine hosts were analyzed. For the capsid protein, a GC content range of 40–44% was observed for strains detected from gorilla, horse, porcine, rabbit, human, and dromedary. Otarine, turkey, and fox

strains have a higher GC content of 47.5, 49.5, and 50.5%, respectively. The RdRp of strains detected from gorilla, horse, dromedary, monkey, human, and bovine has a GC content between 41 and 45%, while a higher GC content was observed for turkey (47%), otarine (48%), and porcine (46%) strains (Supplemental Fig. 2a,b).

The Seg 1 also encodes a hypothetical protein. In the gorilla viromes, the predicted hypothetical proteins were variable in size (66–293 aa) with only 5/15 variants having proteins above 150 aa. A similar observation was done with strains detected from other hosts. For example, strains detected from rabbit and humans also have smaller protein (approximately 100 aa), when compared to strains detected from horses (151aa) (Fig. 3a, b). The hypothetical protein has short repetitive motifs (ExxRxNxxxE) in all the strains, as demonstrated previously in strains detected from human and rabbit [54]. Previous reports have pointed out that short linear motifs can have several functions such as protein degradation, cell signaling, immune response, cell cycle, transcriptional regulation, or translation [54]. However, in that case, the functional role of these motifs remains unknown and further study will be necessary to decipher their biological function.

This study showed that the proportion of intrinsically disordered protein regions (IDPR) in GPbvs was about 15–68% in the hypothetical protein and 14–27% in the capsid protein. IDPR has been previously observed for viral proteins [55] and are important in regulating cellular, viral functions [56]

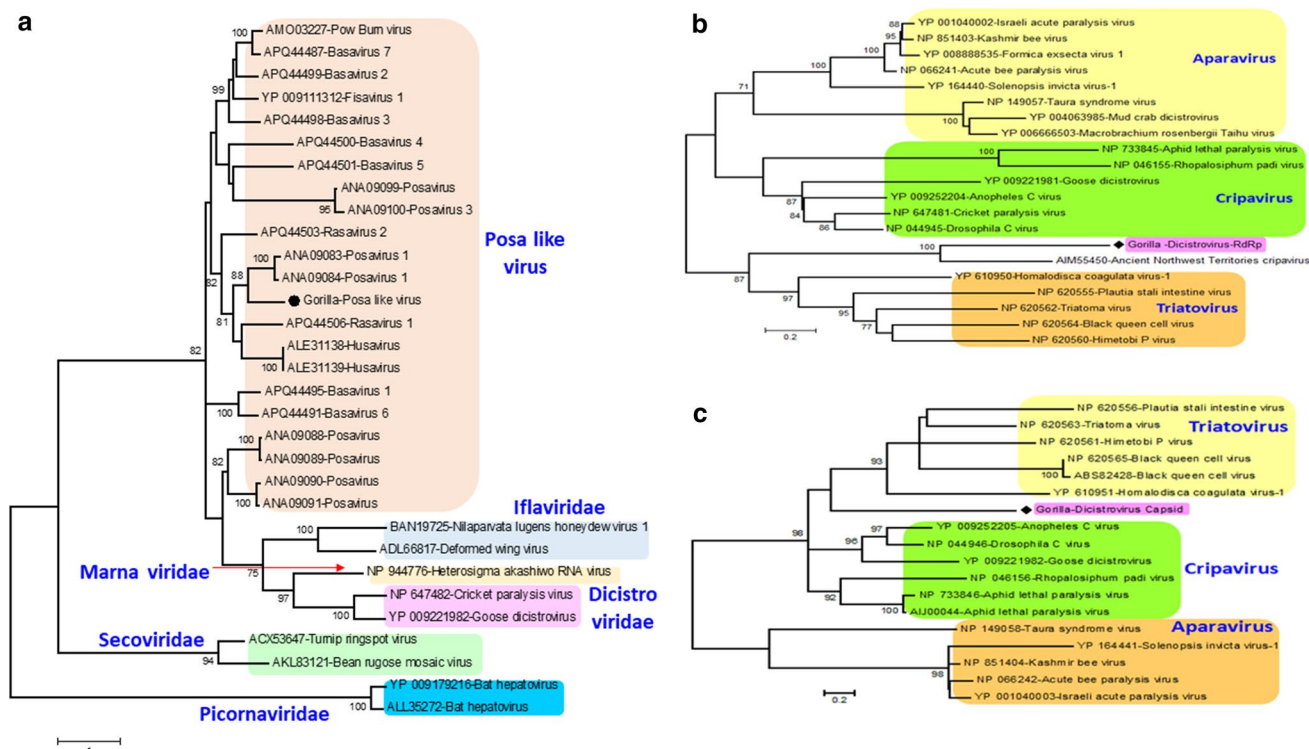


Fig. 6 **a** Phylogenetic analysis of picornavirales (posa-like viruses) sequence including the novel viral sequence from this study. The maximum likelihood tree was constructed based on 314 amino acids with LG + G substitution model with 100 replicates. In the analysis, 31 sequences were included consisting in five recognized families in the *Picornavirales* and unrecognized ‘posa-like viruses’ along with the gorilla stool-associated virus. Strains from known viral families are clustered according to the family. The unclassified ‘posa-like viruses’ formed a separate cluster in the tree. The strain detected from gorillas also branched within this cluster. Each family is highlighted with a shadow and name. **b** Phylogenetic analysis of the RdRp protein of the gorilla stool-associated dicistrovirus and sequences detected

from other hosts. A maximum likelihood (WAG with G + I) tree was constructed based on the analysis of 508 amino acids with 100 replicates. In the tree, representative sequences from all three known genera were included and highlighted in shadow and named. **c** Phylogenetic analysis of the capsid protein (ORF 2) of the gorilla stool-associated dicistrovirus and sequences detected from other hosts. A maximum likelihood tree (620 amino acids) was constructed for the gorilla stool-associated dicistrovirus along with other known species in the family. The tree was constructed using rREV + G + I + F algorithm with 100 replicates. The three recognized genera are highlighted and the strain detected from gorillas is highlighted with bullets and boxes

and help viruses to hijack various pathways of the host cells. The percentage of IDPR in viral proteins is largely variable, from 7.3% for the human coronavirus NL63 to 77.3% for the avian carcinoma virus [55]. In this study, disorder region was particularly detected in the N-terminal region of the capsid proteins. Similar observation was also made by another research team; that is, the C-terminal of nodaviral protein A is a disordered region and it has a relatively high evolutionary rate that can facilitate the rise of novel functions [56]. This first report about detection of disordered protein in PBVs opens new exploratory ways to understand mechanisms driving virus evolution and virus/host interactions.

Although PBVs are detected from many hosts, their evolutionary relationship and potential origin are still unclear. The PBVs were initially classified under *Birnaviridae*, as both are bi-segmented viruses and then reclassified into a new viral family [27]. In 2010, Tang et al. demonstrated

structural similarities between three partitiviruses and *Rabbit picobirnaviruses* (RaPBVs). The RaPBV shares many similarities with partitiviruses in various structural (capsid) properties, including the capsid diameter and thickness [57]. In this study, 63 amino acid sequences (43 from picobirnaviruses and 20 from partitivirus) were analyzed and the presence of a few conserved amino acids and motifs (SGxxxT and GDD) in all the strains (except in radish partitivirus) was identified. This finding supports the earlier hypothesis that the partitivirus might have crossed species barrier from fungi to vertebrates and got adapted or is getting adapted to the host they reside [27]. However, as these two motifs are also observed in dicistroviruses, it may also suggest that these motifs are common to RdRp.

Interspecies transmission (defined as the detection of genetically related viruses between species in one territory) was observed with many viruses, including PBVs in the past. Closely related viruses have been detected between species

from different geographical origins and at different points in time. In Hungary, closely related PBVs were detected both from porcine and humans [58]. Besides, human homolog strains have been detected from different hosts like equine [59], rodents [40], and macaca [1] and this study also found two variants (GPbv-31 and V-39) which were closely related to human strains. An analysis to detect possible recombinant events by two distinct methods suggests that no apparent recombination events were noticed in GPbvs, although recombination events were reported in some PBVs [1]. Finally, as additional virological surveys using metagenomics will be conducted on other healthy and diseased hosts, more (complete) sequences related to PBVs will be characterized. This will further expand our knowledge on the extent of picobirnavirus diversity and role in pathogenesis.

The order *Picornavirales* consists of five recognized families and each family has a unique host range: members of the families *Dicistroviridae* and *Iflaviridae* infect invertebrates, those of *Marnaviridae* infect specific strains of algae, those of *Picornaviridae* mainly infect vertebrates, and those of *Secoviridae* only infect plants [60]. Due to the increased virological surveys and discovery projects, many picorna-like viruses have been detected. Currently, these viruses have been placed under 43 genus and 5 families by the ICTV (International Committee on Taxonomy of Viruses) (<http://www.ictvonline.org/virusTaxonomy.asp>). In 2011, Shan et al. reported the detection of picornavirus sequences from stool samples of porcine, which also share high homology with cDNA sequences derived from *Ascaris suum* (nematode). The viruses were tentatively named as posa (Porcine Stool Associated) viruses [29]. Since then, several studies have reported the identification of ‘posa-like viruses’ from many hosts; human—HUSA virus [30], fish-FISA virus [31], and swine [61] and recently from rat and bats [32]. This group of viruses formed separate clusters in phylogenetic analysis and is proposed as a new viral family in the *Picornavirales*. The present study has identified posa-like virus sequences from gorilla stools (GOSA), which formed a separate branch in the ‘posa-like virus genetic cluster’ with whom it shared less than 37% of identity suggesting this is a novel member in the proposed new viral family.

Currently the family *Dicistroviridae* consists of three genera and all the members are known to infect arthropods [33, 62]. Some of these viruses are pathogenic and have devastating economic consequences [63]. For instance, acute bee paralysis virus, Kashmir bee virus, and Israeli acute paralysis virus infect honey bees, whereas aura syndrome virus and mud crab virus infect shrimps [63]. Although it infects only arthropods, in the recent years, many novel dicistroviruses have been detected from different hosts: goose [64], bat [62], fresh water prawn, panda [65], Mud crab, and Griffin. However, as it is the case here for gorilla stool-associated dicistroviruses, those are likely from a dietary origin (arthropod

consumption) or reflect the presence of intestinal parasites including nematodes [65].

Discovery studies on wild animals offer several advantages over similar studies performed on captive animals as those may be contaminated with human pathogens and/or pathogens acquired from other animals in the zoos. However, working on wildlife is also a difficult task that often limits the number of samples that can be collected and/or the number of individuals that can be sampled. In this study, we performed a global viral metagenomic analysis on stool samples collected from 17 wild gorillas. Although limited by the small number of samples collected and by the absence of virome profiles by individual gorillas (pooling strategy for sequencing), this study identified sequences related to several novel RNA viruses including sequences related to picobirnaviruses that were abundantly covered in the stool samples. The comprehensive genomic sequence analysis explored their genetic and phylogenetic relationship with other picobirnaviruses. Also, the identification of genetic signatures (repetitive amino acids motifs and disordered regions) in the viral protein will facilitate further studies to understand their role in pathogenesis. The identification of novel sequences related to posa-like virus, dicistrovirus, and partitivirus will increase our knowledge in understanding the diversity of these viral families. In all, this study increases our knowledge about viral families associated with gorilla stools of asymptomatic individuals that could serve as baseline datasets for comparative studies in order to identify unusual viruses (potential causative agents) in diseased animals.

Acknowledgement This work was funded through a Mediterranean Infection fellowship to RD, a SANOFI-PASTEUR international prize awarded to CD, and by the French Government under the “Investissements d’avenir” (Investments for the Future) program managed by the National Agency for Research reference Méditerranée Infection 10-IAHU-03.

Author contributions R.D., D.R., and C.D. conceived and designed the experiments; R.D., C.M., and C.R. performed the experiments; J.A., B.D., O.M., and H.-J.P. collected the samples on the field; R.D. analyzed the data; R.D. wrote the article; R.D., P.B., B.D., D.R., and C.D. revised the article.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

Ethical approval The article did not involve in any research with human. The gorilla stools were collected during a mission that was approved by the Ministry of Health (No 208/MSP/CAB.15 of 20 August 2015) and the Forest Economy and Sustainable Development (No 94/MEFDD/CAB/DGACFAP-DTS of 24 August 2015) of the Republic of the Congo.

References

- S.J. Anthony, A. Islam, C. Johnson, I. Navarrete-Macias, E. Liang, K. Jain, P.L. Hitchens, X. Che, A. Soloyvov, A.L. Hicks, R. Ojeda-Flores, C. Zambrana-Torrel, W. Ulrich, M.K. Rostal, A. Petrosov, J. Garcia, N. Haider, N. Wolfe, T. Goldstein, S.S. Morse, M. Rahman, J.H. Epstein, J.K. Mazet, P. Daszak, W.I. Lipkin, *Nat Commun* **6**, 8147 (2015)
- T.F. Ng, W. Zhang, J. Sachsenroder, N.O. Kondov, A.C. da Costa, E. Vega, L.R. Holtz, G. Wu, D. Wang, C.O. Stine, M. Antonio, U.S. Mulvaney, M.O. Muench, X. Deng, K. Ambert-Balay, P. Pothier, J. Vinje, E. Delwart, *Virus Evol.* (2015). <https://doi.org/10.1093/ve/vev017>
- X. Wang, J. Wang, C. Zhou, S. Yang, Q. Shen, W. Zhang, D. Qi, *Gut Pathog* **8**, 53 (2016)
- S. Temmam, B. Davoust, J.M. Berenger, D. Raoult, C. Desnues, *Int. J. Mol. Sci.* **15**, 10377–10397 (2014)
- M. Bermejo, J.D. Rodriguez-Teijeiro, G. Illera, A. Barroso, C. Vila, P.D. Walsh, *Science* **314**, 1564 (2006)
- G. Palacios, L.J. Lowenstine, M.R. Cranfield, K.V. Gilardi, L. Spelman, M. Lukasiak-Braun, J.F. Kinani, A. Mudakikwa, E. Nyirakaragire, A.V. Bussetti, N. Savji, S. Hutchison, M. Egholm, W.I. Lipkin, *Emerg. Infect. Dis.* **17**, 711–713 (2011)
- M.B. Keita, I. Hamad, F. Bittar, *Microb. Pathog.* **77**, 149–154 (2014)
- S. Kondgen, H. Kuhl, P.K. N’Goran, P.D. Walsh, S. Schenk, N. Ernst, R. Biek, P. Formenty, K. Matz-Rensing, B. Schweiger, S. Junglen, H. Ellerbrok, A. Nitsche, T. Briese, W.I. Lipkin, G. Pauli, C. Boesch, F.H. Leendertz, *Curr. Biol.* **18**, 260–264 (2008)
- M.S. Oberste, M.M. Feeroz, K. Maher, W.A. Nix, G.A. Engel, K.M. Hasan, S. Begum, G. Oh, A.H. Chowdhury, M.A. Pallansch, L. Jones-Engel, *J. Virol.* **87**, 558–571 (2013)
- C.N. Nkogue, M. Horie, S. Fujita, M. Ogino, Y. Kobayashi, K. Mizukami, T. Masatani, S. Ezzikouri, A. Matsuu, T. Mizutani, M. Ozawa, O. Yamato, A. Ngomanda, J. Yamagiwa, K. Tsukiyama-Kohara, *Virus Genes* **52**, 671–678 (2016)
- T.A. Seimon, S.H. Olson, K.J. Lee, G. Rosen, A. Ondzie, K. Cameron, P. Reed, S.J. Anthony, D.O. Joly, W.B. Karesh, D. McAloose, W.I. Lipkin, *PLoS ONE* **10**, e0142766 (2015)
- L. Richard, R. Rua, E. Betsem, A. Mouinga-Ondeme, M. Kazanji, E. Leroy, R. Njoum, F. Buseyne, P.V. Afonso, A. Gessain, *J. Virol.* **89**, 12480–12491 (2015)
- L. Richard, A. Mouinga-Ondeme, E. Betsem, C. Filippone, E. Nerrienet, M. Kazanji, A. Gessain, *Clin. Infect. Dis.* **63**, 800–803 (2016)
- S.A. Sadeuh-Mba, M. Bessaud, M.L. Joffret, M.C. Endegue Zanga, J. Balanant, E. Mpoudi Ngole, R. Njoum, J.M. Reynes, F. Delpeyroux, D. Rousset, *PLoS Negl. Trop. Dis.* **8**, e3052 (2014)
- S. Lyons, C. Sharp, M. LeBreton, C.F. Djoko, J.A. Kiyang, F. Lankester, T.G. Bibila, U. Tamoufe, J. Fair, N.D. Wolfe, P. Simmonds, *PLoS ONE* **7**, e33430 (2012)
- F.H. Leendertz, N. Scuda, K.N. Cameron, T. Kidega, K. Zuberbuhler, S.A. Leendertz, E. Couacy-Hymann, C. Boesch, S. Calvignac, B. Ehlers, *J. Virol.* **85**, 916–924 (2011)
- O. Blinkova, J. Victoria, Y. Li, B.F. Keele, C. Sanz, J.B. Ndjongo, M. Peeters, D. Travis, E.V. Lonsdorf, M.L. Wilson, A.E. Pusey, B.H. Hahn, E.L. Delwart, *J. Gen. Virol.* **91**, 74–86 (2010)
- L. Li, A. Kapoor, B. Slikas, O.S. Bamidele, C. Wang, S. Shaukat, M.A. Masroor, M.L. Wilson, J.B. Ndjongo, M. Peeters, N.D. Gross-Camp, M.N. Muller, B.H. Hahn, N.D. Wolfe, H. Triki, J. Bartkus, S.Z. Zaidi, E. Delwart, *J. Virol.* **84**, 1674–1682 (2010)
- N. Dione, S. Khelaifa, B. La Scola, J.C. Lagier, D. Raoult, *Clin. Microbiol. Infect.* **22**(1), 53–58 (2016). <https://doi.org/10.1016/j.cmi.2015.10.032>
- A. Steyer, M. Jevsnik, M. Petrovec, M. Pokorn, S. Grosek, A. Fratnik Steyer, B. Soba, T. Ursic, T. Cerar Kisek, M. Kolenc, M. Trkov, P. Sparl, R. Duraisamy, I.W. Lipkin, S. Terzic, M. Kolnik, T. Mrvic, A. Kapoor, F. Strle, *Pediatr. Infect. Dis. J.* **35**, e262–e270 (2016)
- K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, *Mol. Biol. Evol.* **28**, 2731–2739 (2011)
- K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, *Mol. Biol. Evol.* **30**, 2725–2729 (2013)
- B.M. Muhire, A. Varsani, D.P. Martin, *PLoS ONE* **9**, e108277 (2014)
- K.S. Lole, R.C. Bollinger, R.S. Paranjape, D. Gadkari, S.S. Kulkarni, N.G. Novak, R. Ingersoll, H.W. Sheppard, S.C. Ray, *J. Virol.* **73**, 152–160 (1999)
- G.J. Etherington, J. Dicks, I.N. Roberts, *Bioinformatics* **21**, 278–281 (2005)
- B. Ganesh, G. Masachessi, Z. Mladenova, *Virus Dis.* **25**, 223–238 (2014)
- Y.S. Malik, N. Kumar, K. Sharma, K. Dhama, M.Z. Shabbir, B. Ganesh, N. Kobayashi, K. Banyai, *Biomed. Res. Int.* **2014**, 780752 (2014)
- M.L. Nibert, S.A. Ghabrial, E. Maiss, T. Lesker, E.J. Vainio, D. Jiang, N. Suzuki, *Virus Res.* **188**, 128–141 (2014)
- T. Shan, L. Li, P. Simmonds, C. Wang, A. Moeser, E. Delwart, *J. Virol.* **85**, 11697–11708 (2011)
- B.B. Oude Munnink, M. Cotten, M. Deijis, M.F. Jebbink, M. Bakker, S.M. Farsani, M. Canuti, P. Kellam, L. van der Hoek, *J. Gen. Virol.* **96**, 3440–3443 (2015)
- G. Reuter, P. Pankovics, E. Delwart, A. Boros, *Arch. Virol.* **160**, 565–568 (2015)
- B.B. Oude Munnink, M.V.T. Phan, V. Consortium, P. Simmonds, M.P.G. Koopmans, P. Kellam, L. van der Hoek, M. Cotten, *Virus Evol.* (2017). <https://doi.org/10.1093/ve/vex022>
- B.C. Bonning, W.A. Miller, *Annu. Rev. Entomol.* **55**, 129–150 (2010)
- C.I. Gallimore, H. Appleton, D. Lewis, J. Green, D.W. Brown, *J. Med. Virol.* **45**, 135–140 (1995)
- H.G. Pereira, T.H. Flewett, J.A. Candeias, O.M. Barth, *J. Gen. Virol.* **69**(Pt 11), 2749–2754 (1988)
- B. Ganesh, S.M. Nataraju, K. Rajendran, T. Ramamurthy, S. Kanungo, B. Manna, S. Nagashima, D. Sur, N. Kobayashi, T. Krishnan, *Infect. Genet. Evol.* **10**, 511–516 (2010)
- E. Takiuchi, R. Macedo, A.F. Kunz, J.C. Gallego, J.L. de Mello, R.A. Otonel, A.A. Alfieri, *Virus Res.* **211**, 58–63 (2016)
- T.F. Ng, J.R. Mesquita, M.S. Nascimento, N.O. Kondov, W. Wong, G. Reuter, N.J. Knowles, E. Vega, M.D. Esona, X. Deng, J. Vinje, E. Delwart, *Vet. Microbiol.* **171**, 102–111 (2014)
- M.C. Fregolente, E. de Castro-Dias, S.S. Martins, F.R. Spilki, S.M. Allegretti, M.S. Gatti, *Virus Res.* **143**, 134–136 (2009)
- T.G. Phan, B. Kapusinszky, C. Wang, R.K. Rose, H.L. Lipton, E.L. Delwart, *PLoS Pathog.* **7**, e1002218 (2011)
- P.C. Woo, S.K. Lau, J.L. Teng, A.K. Tsang, M. Joseph, E.Y. Wong, Y. Tang, S. Sivakumar, R. Bai, R. Wernery, U. Wernery, K.Y. Yuen, *Virology* **471–473**, 117–125 (2014)
- G.M. Carruyo, G. Mateu, L.C. Martinez, F.H. Pujol, S.V. Nates, F. Liprandi, J.E. Ludert, *J. Clin. Microbiol.* **46**, 2402–2405 (2008)
- J. Green, C.I. Gallimore, J.P. Clewley, D.W. Brown, *Arch. Virol.* **144**, 2457–2465 (1999)
- B. Ganesh, K. Banyai, G. Masachessi, Z. Mladenova, S. Nagashima, S. Ghosh, S.M. Nataraju, M. Pativada, R. Kumar, N. Kobayashi, *Vet. Res.* **42**, 52 (2011)
- H. Verma, S.K. Mor, J. Erber, S.M. Goyal, *Infect. Genet. Evol.* **30**, 134–139 (2015)
- R. Bodewes, J. van der Giessen, B.L. Haagmans, A.D. Osterhaus, S.L. Smits, *J. Virol.* **87**, 7758–7764 (2013)

47. L. Gillman, A.M. Sanchez, J. Arbiza, *Intervirology* **56**, 46–49 (2013)
48. Y. Wang, X. Tu, C. Humphrey, H. McClure, X. Jiang, C. Qin, R.I. Glass, B. Jiang, *J. Med. Primatol.* **36**, 101–107 (2007)
49. G. Masachessi, B. Ganesh, L.C. Martinez, M.O. Giordano, P.A. Barril, M.B. Isa, G.V. Pavan, C.A. Mateos, S.V. Nates, *Infect. Genet. Evol.* **29**, 196–202 (2015)
50. P.C. Woo, S.K. Lau, R. Bai, J.L. Teng, P. Lee, P. Martelli, S.W. Hui, K.Y. Yuen, *J. Virol.* **86**, 6377–6378 (2012)
51. R. Bodewes, A. Ruiz-Gonzalez, C.M. Schapendonk, J.M. van den Brand, A.D. Osterhaus, S.L. Smits, *Virol. J.* **11**, 89 (2014)
52. G. Masachessi, L.C. Martinez, B. Ganesh, M.O. Giordano, P.A. Barril, M.B. Isa, A. Ibars, J.V. Pavan, S.V. Nates, *Arch. Virol.* **157**, 2075–2082 (2012)
53. E.M. Symonds, D.W. Griffin, M. Breitbart, *Appl. Environ. Microbiol.* **75**, 1402–1409 (2009)
54. B. Da Costa, S. Duquerroy, B. Tarus, B. Delmas, *Virus Res.* **158**, 251–256 (2011)
55. B. Xue, A.K. Dunker, V.N. Uversky, *J. Biomol. Struct. Dyn.* **30**, 137–149 (2012)
56. L. Gitlin, T. Hagai, A. LaBarbera, M. Solovey, R. Andino, *PLoS Pathog.* **10**, e1004529 (2014)
57. J. Tang, W.F. Ochoa, H. Li, W.M. Havens, M.L. Nibert, S.A. Ghabrial, T.S. Baker, *J. Struct. Biol.* **172**, 363–371 (2010)
58. K. Banyai, V. Martella, A. Bogdan, P. Forgach, F. Jakab, E. Meleg, H. Biro, B. Meleg, G. Szucs, *J. Gen. Virol.* **89**, 534–539 (2008)
59. G.F. Browning, R.M. Chalmers, D.R. Snodgrass, R.M. Batt, C.A. Hart, S.E. Ormarod, D. Leadon, S.J. Stoneham, P.D. Rossdale, *Equine Vet. J.* **23**, 405–409 (1991)
60. O. Le Gall, P. Christian, C.M. Fauquet, A.M. King, N.J. Knowles, N. Nakashima, G. Stanway, A.E. Gorbalenya, *Arch. Virol.* **153**, 715–727 (2008)
61. B.M. Hause, R.A. Hesse, G.A. Anderson, *Virus Genes* **51**, 144–147 (2015)
62. G. Reuter, P. Pankovics, Z. Gyongyi, E. Delwart, A. Boros, *Arch. Virol.* **159**, 3453–3456 (2014)
63. S.M. Valles, Y. Chen, A.E. Firth, D.M. Guerin, Y. Hashimoto, S. Herrero, J.R. de Miranda, E. Ryabov, *C. Ictv Report, J. Gen. Virol.* **98**, 355–356 (2017)
64. A.L. Greninger, K.R. Jerome, *Genome Announc.* **4**, 68 (2016)
65. W. Zhang, S. Yang, T. Shan, R. Hou, Z. Liu, W. Li, L. Guo, Y. Wang, P. Chen, X. Wang, F. Feng, H. Wang, C. Chen, Q. Shen, C. Zhou, X. Hua, L. Cui, X. Deng, Z. Zhang, D. Qi, E. Delwart, *Microbiome* **5**, 90 (2017)