Multiple divergent picobirnaviruses with functional prokaryotic Shine- Dalgarno ribosome binding sites present in cloacal sample of a diarrheic chicken

Ákos Boros[a,b], Beáta Polgár[b], Péter Pankovics[a,b], Hajnalka Fenyvesi[a,b], Péter Engelmann[c], Tung Gia Phan[d,e], Eric Delwart[d,e], Gábor Reuter[b,□]

a Regional Laboratory of Virology, National Reference Laboratory of Gastroenteric Viruses, ÁNTSZ Regional Institute of State Public Health Service, Pécs, Hungary

b Department of Medical Microbiology and Immunology, Medical School, University of Pécs, Pécs, Hungary

c Department of Immunology and Biotechnology, Clinical Center, Medical School, University of Pécs, Pécs, Hungary

d Blood Systems Research Institute, San Francisco, CA, USA

e University of California, San Francisco, CA, USA

□ Correspondence to: Department of Medical Microbiology and Immunology Medical School, University of Pécs, Szigeti út 12, H-7624 Pécs, Hungary.
E-mail address: reuter.gabor@pte.hu (G. Reuter).

Abstract

Picobirnaviruses (PBVs) of family *Picobirnaviridae* have bisegmented (S1 and S2 segments), double-stranded RNA genomes. In this study a total of N = 12 complete chicken PBVs (ChPBV) segments (N = 5 of S1 and N = 7 of S2, Acc. Nos.: MH425579-90) were determined using viral metagenomic and RT-PCR techniques from a single cloacal sample of a diarrheic chicken. The identified ChPBV segments are unrelated to each other and distant from all of the currently known PBVs. *In silico* sequence analyses revealed the presence of conserved prokaryotic Shine-Dalgarno-like (SD-like) sequences upstream of the three presumed open reading frames (ORFs) of the S1 and a single presumed ORF of the S2 segments. According to the results of expression analyses in *E. coli* using 6xHis-tagged recombinant ChPBV segment 1 construct and Western blot these SD-like sequences are functional in vivo suggesting that S1 of study PBVs can contain three ORFs and supporting the bacteriophage-nature of PBVs.

1. Introduction

Picobirnaviruses (PBVs) of family *Picobirnaviridae* are small (Ø 33–41 nm), non-enveloped viruses with bisegmented, double-stranded RNA (dsRNA) genome (Pereira et al., 1988; Duquerroy et al., 2009; Delmas, 2011). Matured PBV virions contain two functionally unrelated, linear dsRNA segments (called as segment 1 and 2) in equimolar concentration (Pereira et al., 1988; Green at al, 1999; Delmas, 2011) although PBVs with unsegmented genomes were also described recently (Li et al., 2015; Luo et al., 2018). The length of these linear genomic segments are ranged between 2.2 and 2.7 kbp (segment 1) and 1.2–1.9 kbp (segment 2) (Malik et al., 2014).

The segment 1 (S1) could contain multiple, sometimes overlapping ORFs. The S1-ORF1 is *in silico* identifiable at the 5′ end of the segment 1 of certain rabbit and human PBVs (Green at al, 1999; Ng et al., 2014; Delmas, 2011), although due to the unknown function(s) and extreme variations in sequence lengths and identities the functionality or even the presence of this ORF is unclear (Delmas, 2011; Da Costa et al., 2011). The presence of this ORF has not even been mentioned in the PBV-related studies published recently by other groups (Li et al., 2015; Duraisamy et al., 2018; Luo et al., 2018). The S1-ORF2 sequences of different PBVs encode hydrophilic peptides with unknown functions. These peptides predominantly contain conserved and four to ten times repeated sequence motifs of ExxRxNxxxE (x = variable amino acid) which is one of the main characteristic genomic features of PBVs (Da Costa et al., 2011). The third ORFs (S1-ORF3) encode the capsid pre-cursors of PBVs (Delmas, 2011; Ganesh et al., 2012; Malik et al., 2014). The segment 2 (S2) of PBVs contains a single open reading frame (S2-ORF1) encoding the RNA-dependent RNA polymerase (RdRp) enzyme (Ganesh et al., 2012; Malik et al., 2014). Based on the analysis of short (≈200–370 bp) partial RdRp sequences of segment 2 generated by the widely used screening oligonucleotide primer sets two main genogroups (GI and GII) of PBVs were distinguished (Carruyo et al., 2008; Rosen et al., 2000; Ganesh et al., 2012). However, using viral metagenomic techniques several non-GI/GII PBVs have been recently discovered (Woo et al., 2014; Li et al., 2015; Duraisamy et al., 2018). The genomic diversity of PBVs identified mainly from fecal samples of healthy and diarrheic cases, less frequently in respiratory or serum samples are remarkably wide (Silva et al., 2014; Duraisamy et al., 2018; Woo et al., 2014; Anthony et al., 2015; Luo et al., 2018; Ganesh et al., 2014; Smits et al., 2011; Li et al., 2015). Picobirnaviruses were detected in wide spectrum of vertebrate species including poultry as well as different environmental samples (for details see reviews of Malik et al., 2014, Ganesh et al., 2014). Due to the lack of permissive vertebrate cell lines to culture PBVs and lack of vertebrate virus families phylogenetically related to PBVs (the closest viruses to PBVs are partitiviruses of fungal origin and atypical PBVs of *Cryptosporidium parvum*) vertebrates as natural hosts of PBVs could be questionable (Green et al., 1999).

While the overall enteric viral diversity including PBV diversity of chickens identified from cloacal

pools has been investigated before (Silva et al., 2014; Day and Zsak, 2013; Day et al., 2015) until now little is known about the enteric viral composition and diversity of individual diarrheic chicken.

In this study the total of 12 novel, full length sequences (N = 5 of segment 1 and N = 7 of segment 2) of chicken PBVs (ChPBVs) were determined using metagenomic and RT-PCR techniques from a single cloacal sample of a diarrheic chicken which was previously reported to contain co-infections of eight picornaviruses (family *Picornaviridae*) (Boros et al., 2016). Furthermore, sequence analysis of identified ChPBVs revealed the presence of conserved sequence motifs prior to all ORFs closely resemble to the ribosome-binding motifs of prokaryotes (Shine-Dalgarno/SD/sequences) as recently described in other PBV segments (Shine and Dalgarno, 1975; Krishnamurthy and Wang, 2018). According to the results of expression analyses in *E. coli* using 6xHis- tagged recombinant ChPBV segment 1 construct and Western blot assay these SD-like sequences could be functional in vivo and also implying that segment 1 of PBVs could contain three functional ORFs. These results suggest that - in accordance with recent studies published by the time of the manuscript preparation (Krishnamurthy and Wang, 2018; Yinda et al., 2018) - PBVs could be in fact prokaryotic rather than eukaryotic viruses.

2. Materials and methods

2.1. Sample preparation, next-generation sequencing and viral metagenomics

The same viral metagenomic sequence data pool which was generated from the next-generation sequencing (NGS) of single cloacal specimen (sample ID Pf-Chk-1) was used in this study for the detection of picobirnaviruses as used for the analysis of picornaviruses (Boros et al., 2016). Therefore, the methods of sample preparation, NGS (MiSeq Il- lumina platform) and a viral metagenomic analysis including BLASTx analyses were described previously (Boros et al., 2016).

2.2. Segment acquisitions by conventional and adapter-ligated RT-PCR

The contigs and singletons related to picobirnaviruses (by BLASTx analysis) were used as targets for specific oligonucleotide primer design. Segment-specific primers were designed to amplify the over-lapping genome regions by different conventional and adapter-ligated RT-PCR reactions described previously (Wakuda et al., 2005) with minor modifications. Briefly, the total RNA used for the adapter-ligation reaction was extracted by TRI reagent (MRC) according to the manufacturer's instructions. The total RNA was denatured at 95 °C for 2 min and then immediately put on ice for 5 min. For the RNA-adapter ligation reactions Adapter A (Phosphorylated-5′ CCC TCG AGT ACT AAC TAG TTA ACT GAT CAC CTC TAG ACC TTT 3′-NH₂-blocked) and T4 RNA ligase enzyme (10U/ reaction) (ThermoFisher Scientific, USA) were used and the reaction mixture was incubated for 1 h at 37 °C. The adapter-ligated RNA was purified using QIAamp Viral RNA Mini Kit (Qiagen). The reverse transcription reaction on the adapter-ligated total RNA was carried out using Adapter B (5′ AAA GGT CTA GAG GTG ATC AGT TAA

CTA GTT AGT ACT CGA GGG 3′) as reverse primer and Maxima H Minus Reverse Transcriptase (ThermoFisher Scientific, USA) according to the manufacturer's instructions and incubated for 1 h at 60°C. Two-step PCR (first and second PCR) was carried out using DreamTaq polymerase (ThermoFisher Scientific, USA) in the otherwise same conditions as described previously (Wakuda et al., 2005). The generated RT-PCR products were sequenced in both directions using dye-terminator sequencing method and run on an automated sequencer ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Stafford, Texas, USA). The sequences of the RT-PCR products were used for the verification of the singletons and contigs and for the analysis of the unknown genome parts not covered by metagenomic sequences including the 5′ and 3′ terminal ends. The oligonucleotide primers used for the acquisition of the segments were available on request.

2.3. Relative quantification of ChPBV segments by RT-qPCR

The total RNA extract was denatured at 95 °C for 2 min and then immediately put on ice for 5 min. The denatured RNA was reverse transcribed using random hexamer (Promega) (500 ng/reaction) Maxima H Minus Reverse Transcriptase (ThermoFisher Scientific, USA) and incubated at 25 °C for 10 min followed by 42 °C for 60 min 1–1 μl of this cDNA was used as template for the downstream qPCR reactions. ChPBV segment quantification relied on segment-specific primer pairs with similar Tm ranged between 63 and 65 °C which were generating RT-qPCR product ranged between 121 and 141 bp. Maxima SYBR Green qPCR Master Mix (ThermoFisher Scientific, USA), segment specific primer pairs with a final concentration of 0.3–0.3 μM and 1–1 μl of cDNA as template was used in the individual qPCR reactions. The quantification steps of all of the qPCR reactions were followed by a dissociation assay (melting point analysis) and both of them were run on a LightCycler 2.0 thermal cycler (Roche, Germany). The qPCR analyses were included three technical repeats. The oligonucleotide primers used for the qPCR reactions were available on request.

2.4. Sequence and phylogenetic analyses of ChPBV segments

The nucleotide (nt) and amino acid (aa) sequences of picobirnavirus segments were aligned using the MUSCLE algorithm. The pairwise nt and aa identity calculations of the aligned sequences were performed using Sequence Identity And Similarity (SIAS) web tool (http://imed.med.ucm.es/Tools/sias.html). Phylogenetic trees with Neighbor-Joining method and Jones–Taylor–Thornton matrix-based model were constructed using MEGA software ver. 6.06 (Tamura et al., 2013). Bootstrap values were set to 1000 replicates and only likelihood percentages of > 50% were indicated in the trees. The full-length sequences of the identified ChPBV segments were submitted to GenBank with the accession numbers of MH425579 - MH425583 of Segment 1 sequences and

MH425584 - MH425590 of segment 2 sequences.

2.5. Cultivation of enteric bacteria and analysis of potential virus growth

The bacterial cultivation experiments were conducted by the description of Kim and co-workers with minor modification (Kim et al., 2011). 10–10 µl of 3% (w/v) cloacal sample suspension diluted with 0.1 M phosphate-buffered saline (PBS) was inoculated into 10–10 ml of brain heart infusion (BHI) broth in duplicate. The in vitro cultures were incubated at 37 °C for up to two weeks in aerobic and anaerobic conditions. 150–150 µl from the inoculated media were collected in every two days used for RNA extraction by the same TRI reagent (MRC) and SYBR green-based RT-qPCR with the same segment-specific oligonucleotide primers as described above.

2.6. Generation of His-tagged dsDNA construct

For the generation of a recombinant 1676-bp-long dsDNA construct which encodes three artificially 6xHis-tagged viral peptides of an identified ChPBV S1 sequence (ChPBV-S1-Ctg289/2013-HUN) the same cDNA sample was used as template which was described in the *Relative quantification of ChPBV segments by RT-qPCR* section. Table 1 summarizes the steps of PCR and ligation reactions. For the PCR reactions high-fidelity *Pfu* polymerase (ThermoFisher Scientific, USA) and 5′end modified oligonucleotide primers (Table 1) targeting the sequence of ChPBV-S1-Ctg289/2013-HUN were used. The PCR products were purified before all ligation reactions using GeneJET PCR Purification Kit (ThermoFisher Scientific, USA). In order to successfully separate all of the three potential viral peptides in a single polyacrylamide gel, only the 5′ 918 nt part of the S1-ORF3 was used in the construct (Table 1). For the DNA ligation reactions T4 DNA Ligase enzyme (ThermoFisher Scientific, USA) was used according to the manufacturer's instructions and the reaction mixture was incubated at 22 °C for 1 h. The purified 1676-nt-long recombinant dsDNA construct containing a T7-promoter sequence at the 5' end and all of the three *in silico* identifiable viral ORFs of ChPBV-S1-Ctg289/2013-HUN with an artificially added C-terminal 6xHis-encoding tag (Fig. 4A) was used as an insert of the following cloning and expression analyses.

Table 1. Steps and main features (templates and oligonucleotide primers) of PCR and ligation reactions used for the generation of final, 1676 bp-long dsDNA construct. Lowercase letters in the primer sequences represents the 6xHis-tag, underlined sequences represents the T7 promoter sequence. 5′ Phos: 5′ phosphorylated primer.

Reaction type and no.	Template	Primer name	Primer sequence (5′ - 3′)	Product length
PCR-1	cDNA	289-ORF1-T7-F	taatacgactcactatagggGTAAATTTAAACACATCTTAATGAA	231 bp
		289-ORF1-6xHis-R/5Phos/	/5'Phos/ttatcaatgatgatgatgatgatgGTCATTATATACCTCCTTTCTG	
PCR-2	cDNA	289-ORF2-F/5Phos/	/5'Phos/TAGTCAGAAAGGAGGTATATAATG	489 bp
		289-ORF2-6xHis-R	ttatcaatgatgatgatgatgatgTTTAATACTACCTGCTGCCATA	
PCR-3	cDNA	289-ORF3-F/5Phos/	/5'Phos/TAGGAGGAATAAATATGAATAAGAAA	956 bp
		289-ORF3-6xHis-R	ttatcaatgatgatgatgatgatgTTCAAGATTTTCAGCTATGGCA	
Ligation-1	PCR products of PCR-1 and -2	–	–	720 bp
PCR-4	Product of Ligation-1	289-ORF1-T7-F	taatacgactcactatagggGTAAATTTAAACACATCTTAATGAA	956 bp
		289-ORF2-6xHis-R /5Phos/	/5'Phos/ttatcaatgatgatgatgatgatgTTTAATACTACCTGCTGCCATA	
Ligation-2	PCR products of PCR-3 and -4	–	–	1676 bp
PCR-5	Product of Ligation-2	289-ORF1-T7-F	taatacgactcactatagggGTAAATTTAAACACATCTTAATGAA	1676 bp
		289-ORF3-6xHis-R /5Phos/	/5'Phos/ttatcaatgatgatgatgatgatgTTCAAGATTTTCAGCTATGGCA	

2.7. Cloning and transformation of Escherichia coli

The purified 1676-nt-long recombinant dsDNA construct was cloned into a pJET1.2/blunt vector using the CloneJET PCR Cloning Kit (ThermoFisher Scientific, USA) according to the manufacturer's instructions (ctg289+). For negative control (Neg.K) self-ligated pJET1.2/blunt plasmids were used. After cloning the ctg289+ and Neg.K plasmids were transformed into chemically competent *E. coli* BL21(DE3)pLysS strain (Merck Darmstadt, Germany) using TransformAid Bacterial Transformation Kit (ThermoFisher Scientific, USA). The successfully transformed colonies were selected on Luria-Bertani (LB) agar plates supplemented with 100 µg/ml ampicillin (LB+Amp) by overnight (ON) incubation at 37 °C. Single colonies were then inoculated into 10 ml of LB+Amp media and incubated ON at 37 °C. Plasmids were isolated from the ON cultures using Zyppy™ Plasmid Miniprep Kit (Zymo Research, Irvine, USA). The integrity of the insert was verified by sequencing in both directions of the isolated plasmids. For expression of the 6xHis-tagged viral peptides 500–500 µl of the ON cultures (in 1:100 inoculum:media rate) were inoculated into 50–50 ml of freshly prepared LB+Amp media and incubated in an orbital shaker at 37ºC until the optical density measured at a wavelength of 600 nm (OD_{600}) reached 0.5. 1–1 ml of inoculated media was collected from the cultures before the Isopropyl β-D-1-thiogalactopyr-anoside (IPTG) induction. The cultures then rapidly cooled down to 20ºC and recombinant protein expression was induced with a final concentration of 0.1 mM IPTG at 20 °C, 200 rpm for 17 h in an orbital shaker. When the incubation period was over the cells were harvested with centrifugation at 4.000 rpm for 30 min at 4 °C. All of the following procedures were performed at 4 °C. The obtained pellet was resuspended in a lysis buffer containing 20 mM PBS (pH 7.6), 300 mM NaCl, 5 mM imidazole, 0.1% v/v Triton X-100 (Sigma-Aldrich, Germany), 1/10 vol Lysozyme (10 mg/ml stock) and one Protease Inhibitor mini tablet (ThermoFisher Scientific, USA). The resuspended pellet was incubated in an orbital shaker for 30 min on ice then soni- cated on ice, for 30 s/cycle with 0.5 s pulse at 30% amplitude repeated for 5 times. After sonication the bacterial lysates were centrifuged at 13.000 rpm for 10 min at 4 °C and the pellet and supernatant fractions were used for denaturating sodium dodecyl-sulfate-polyacrylamide gelelectrophoresis (SDS-PAGE) and Western blot (WB).

2.8. Denaturating polyacrylamide gel electrophoresis and Western blot

Soluble and pellet fractions of the recombinant construct-transfected (ctg289+) and negative control (Neg.K) *E. coli* lysates as well as 20x diluted 6xHis-tagged recombinant peptide with known molecular mass (25.5 kDa) as a positive control (His+K) were separated on 15% SDS-PAGE and electrotransferred to 0.2 µM pore size nitrocellulose membranes at 54 mA 4 °C ON. The efficacy of the protein transfer was verified by Ponceau-S staining. Then membranes were air-died and nonspecific binding to nitrocellulose membranes was blocked with 5% BSA in 25 mM Tris-base-saline + 0.05% Tween 20 (TBST) buffer (pH 7.4) at RT for 2 h. The presence of expressed His-tagged recombinant constructs were detected by 1:1000

diluted 6x-His Tag-specific monoclonal antibody (Thermo-Fisher Scientific, USA) in 1% BSA-TBST at 4 °C ON. After 4 × 15 min washing with TBST buffer, membranes were incubated with horseradish peroxidase (HRP)-conjugated polyclonal antimouse-IgG antibody (DAKO, Glosup, Denmark) at a dilution of 1:2000 in 1% BSA-TBST at RT for an hour. After 4x washing with TBST chemiluminescent detection of the immunoreactive bands was performed with Western Lightning ECL-Plus reagent kit (Perkin Elmer Life Sciences, Waltham, MA, USA). Bands were visualized on X-ray films and further analyzed with Vilber-Lourmat Bio-Profil Version 97 gel documentation system using Bio-Capt Version 12.6 software.

3. Results

3.1. Identification and categorization of ChPBV segments

The total of 13,016 unique viral reads was generated using next-generation sequencing and viral metagenomics from a single cloacal specimen (sample ID: Pf-Chk-1). According to BLASTx analyses N = 516 sequence reads are possibly originated from picobirnaviruses (Fig. S1). For the detailed analysis of metagenomic sequences see the previous report of the same sample (Boros et al., 2016). Total of 6 contigs (Ctg 289, 626, 725, 811, 883 and 1042) were *in silico* generated from the PBV sequences. Using specific oligonucleotide primer sets designed to the contigs and the individual reads (singletons) of PBVs applied in conventional RT-PCR and adapter-ligated 5′/3′ RACE PCR reactions the full length sequences of chicken PBV (ChPBV) segments (N = 12) were determined (Table 2). Separation attempts of the identified dsRNA segments using polyacrylamide gelelectrophoresis with silver staining (SS-PAGE) method described previously (Herring et al., 1982) showed negative results (data not shown). From the N = 12 sequences N = 5 are segment 1 (S1) sequences and N = 7 are segment 2 (S2) sequences (Table 2, Fig. 1). Based on the results BLASTp analyses and pairwise sequence comparisons the identified ChPBV segments were distantly related to each other or to the known PBV sequences as the closes match found in the GenBank (Table S1; Table 3). Due to practical reasons only the short names (indicate in bold in Tables 2, 3) of the identified segments were used in the manuscript.

Table 2. Genomic features including the lengths of the different genome parts and terminal sequences of ChPBV S1 and S2 segments identified in this study. Identical nucleotides of the 3′ end of ChPBV-S2-ctg1042/2013-HUN and ChPBV-S1-ctg811/2013-HUN are in bold and underlined. Short names of the segments found in the manuscript are in bold.

| Segment name	Accession no.	Total length	5′ end	5'UTR	ORF1	ORF2	ORF3	3′ UTR	3′ end
ChPBV-S1-ctg289/2013-HUN	MH425579	2378 nt	GUAAAuuua...	43 nt	147 nt (48 aa)	447 nt (148 aa)	1710 nt (569 aa)	28 nt	...cccguaucc
ChPBV-S1-ctg626/2013-HUN	MH425580	2536 nt	GUAAAuuuu...	43 nt	189 nt (62 aa)	654 nt (217 aa)	1653 nt (550 aa)	35 nt	...ccacuuacc
ChPBV-S1-ctg725/2013-HUN	MH425581	2798 nt	GUAAAauua...	41 nt	180 nt (59aa)	834 nt (277 aa)	1698 nt (565 aa)	82 nt	...cuacaagcg
ChPBV-S1-ctg811/2013-HUN	MH425582	1989 nt	GUAAAuuaa...	45 nt	168 nt (55 aa)	189 nt (62 aa)	1530 nt (509 aa)	45 nt	...cccuacugc
ChPBV-S1-ctg883/2013-HUN	MH425583	2489 nt	GUAAAuaua...	52 nt	141 nt (46 aa)	588 nt (195 aa)	1620 nt (539 aa)	48 nt	...cgcccuugc
ChPBV-S2-ctg1042/2013-HUN	MH425584	1671 nt	GUAAAugaa...	42 nt	1599 nt (532 aa)	–	–	30 nt	...cccuaguuc
ChPBV-S2-Nov1/2013-HUN	MH425585	1677 nt	GUAAAgauu...	47 nt	1599 nt (532 aa)	–	–	31 nt	...accucugug
ChPBV-S2-Nov2/2013-HUN	MH425586	1652 nt	GUAAAauaa...	52 nt	1569 nt (522 aa)	–	–	31 nt	...ucucuuucu
ChPBV-S2-Nov3/2013-HUN	MH425587	1667 nt	GUAAAacuu...	50 nt	1569 nt (522 aa)	–	–	48 nt	...ugguugggu
ChPBV-S2-Nov4/2013-HUN	MH425588	1772 nt	GUAAAauu...	46 nt	1686 nt (561 aa)	–	–	40 nt	...cgggacacc
ChPBV-S2-Nov5/2013-HUN	MH425589	1616 nt	GUAAAuuuc...	75 nt	1521 nt (506 aa)	–	–	20 nt	...cgaacugaa
ChPBV-S2-Nov6/2013-HUN	MH425590	1664 nt	GUAAAauuu...	48 nt	1599 nt (532 aa)	–	–	17 nt	...acgguggca
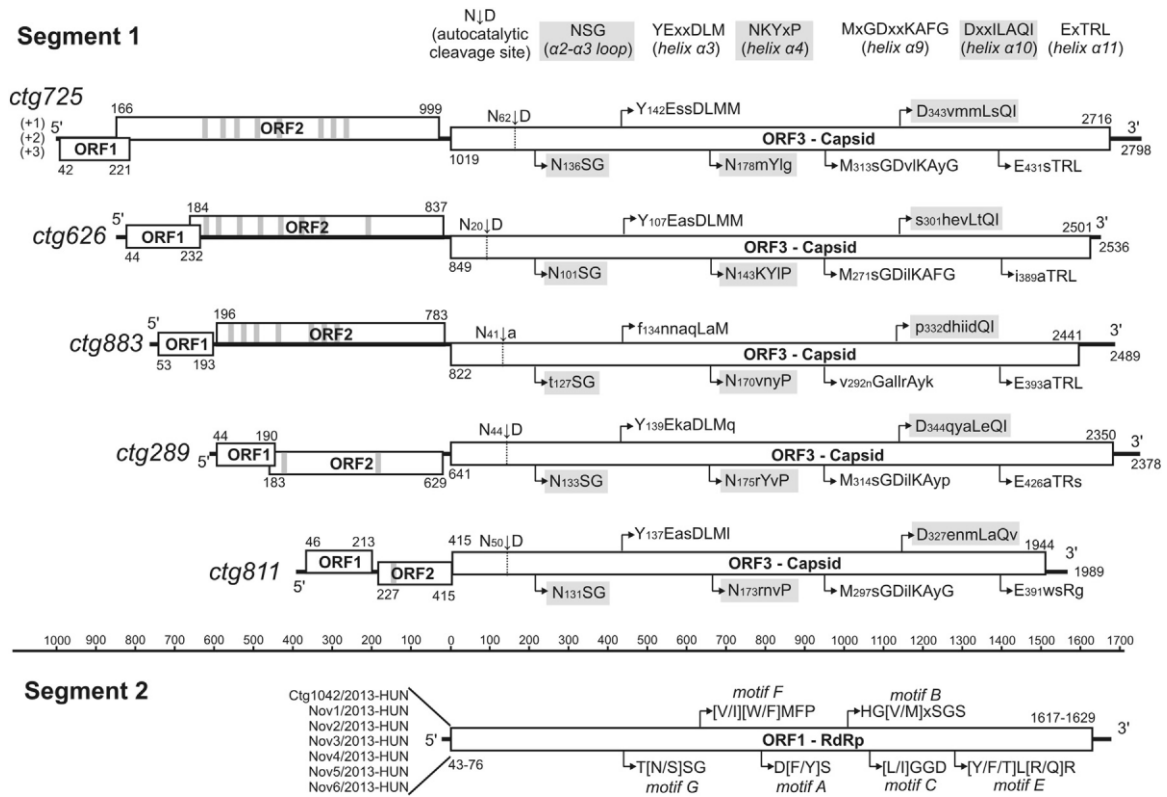
Fig. 1. Schematic genome maps of segment 1 (N = 5) and segment 2 (N = 7) of picobirnaviruses (PBVs) identified in this study. The first and last nucleotide (nt) positions (or from - to positions in case of segment 2 map) of the given open reading frame (ORF) cassettes are marked at borders of every cassette. (+1, +2, +3): 1st, 2nd or 3rd reading frame. Vertical grey lanes in ORF2 cassettes represent the variable localizations of ExxRxNxxxE (x = variable amino acids) repeating motifs (Da Costa et al., 2011). Aa motifs (with positions of the first aa) which could be correspond to the conserved motifs of PBV capsids (Duquerroy et al., 2009; top of the picture) are marked above or below the capsid-encoding ORF3 cassettes (uppercase: conserved, lower case letters: variable aas). Numbers above the ruler indicate nt positions counted from the junction of ORF2 and ORF3. Note that due to similar genomic architecture the of the identified segment 2 sequences only one re- presentative map was depicted. The six of the seven polymerase motifs identified in the RNA-dependent RNA polymerase (RdRp) of human PBV strain Hy005102 (Collier at al, 2016) are labeled sequentially above and below the RdRp cassette as motif G, F, A, B, C, and E (motif D is not reliably identifiable).

3.2. Sequence analyses of complete ChPBV segments

The lengths of S1 sequences ranged between 1989 bp and 2798 bp (Table 2, Fig. 1). There are several potential overlapping ORFs identifiable *in silico* using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) in the S1 study sequences (data not shown), although, only three of them were consistently identifiable in all S1 study sequences and were designated as S1-ORF1-3 (Fig. 1). The short S1-ORF1 located 5' ends of S1 sequences and the downstream S1-ORF2 usually partially overlapped and are therefore present in different reading frames (Fig. 1). The S1-ORF1s ranged between 141 and 189 nucleotide (nt) (Table 2, Fig. 1). Short peptides encoded by S1-ORF1s have no significant sequence hits found in the GenBank by BLASTp and are also unrelated to each other (Table 3,

Table S1). The second S1-ORFs encode hydrophilic peptides ranging between 62 amino acid (aa) and 277 aa which contain the conserved motif of ExxRxNxxxE (x = variable aa) repeated up to eight times (Table 2, Fig. 1). Size variations of the S1- ORFs are mainly responsible for the size variations of the entire segments (Fig. 1). The S1-ORF3s are the longest identifiable ORFs and can encode peptides ranging from 509 to 569 aa (Table 2). Based on the presence of conserved sequence motifs identified in the capsid sequences of other PBVs (Duquerroy et al., 2009) the peptides of S1- ORF3s may be capsid pro-peptides of the identified ChPBVs (Fig. 1).

The lengths of S2 sequences ranged between 1616 bp and 1772 bp (Table 2, Fig., 1). Based on the presence of conserved polymerase motifs identified in the RdRp of human PBVs (Collier et al., 2016) the single longest S1-ORF1 predictable in all (N = 7) S2 sequences ranged be- tween 1521 and 1686 nt (Table 2) encodes the RdRp-s of the ChPBVs (Fig. 1).

Table 3 Results of BLASTp searches using putative peptides encoded by the different open reading frames (ORF) of ChPBV segments as query sequences. Seg. Segment; Acc. No.: Accession number. Short names of the segments found in the manuscript are in bold.

Segment name	Seg.	ORF	Closest match - definition	Organism	Acc. no.	Identity	Query cover	E-value
ChPBV-S1-ctg289/2013-HUN	1	ORF1	tRNA(Asn) ligase	*Euryarchaeota archaeon*	OUV58115	44%	93%	5.3
	1	ORF2	hypothetical protein	Gorilla picobirnavirus	AVD97113	33%	69%	64.0
	1	ORF3	capsid protein	Macaque picobirnavirus 1	AVD54024	28%	99%	7e−58
ChPBV-S1-ctg626/2013-HUN	1	ORF1	LPXTG cell wall anchor domain	*Streptococcus suis*	WP_024532374	28%	98%	35.0
	1	ORF2	hypothetical protein [ORF1]	Macaque picobirnavirus 2	AVD54025	62%	99%	1e−86
	1	ORF3	capsid protein	Macaque picobirnavirus 2	AVD54026	60%	88%	0.0
ChPBV-S1-ctg725/2013-HUN	1	ORF1	tRNA (Arg) ligase	*Candidatus Magasanikbacteria*	PJA89872	27%	93%	5.2
	1	ORF2	hypothetical protein [ORF1]	Otarine picobirnavirus	AMP18928	30%	72%	9e−09
	1	ORF3	capsid protein	Porcine picobirnavirus	ASM93458	30%	98%	1e−54
ChPBV-S1-ctg811/2013-HUN	1	ORF1	ninF protein	*Escherichia coli*	WP_048970988	33%	100%	4.9
	1	ORF2	hypothetical protein	Turkey picobirnavirus	AHZ46148	39%	79%	7.5
	1	ORF3	capsid protein	Dromedary picobirnavirus	AIY31275	97%	32%	1e−64
ChPBV-S1-ctg883/2013-HUN	1	ORF1	DUF2764 domain-containing protein	*Chlamydia abortus*	WP_011097238	31%	97%	71.0
	1	ORF2	hypothetical protein [ORF1]	Macaque picobirnavirus 9	AVD54039	74%	100%	3e−98
	1	ORF3	capsid protein	Human picobirnavirus	Q50LE5	27%	45%	1e−18
ChPBV-S2-ctg1042/2013-HUN	2	ORF1	RNA-dependent RNA polymerase	Gorilla picobirnavirus	AVA30702	76%	99%	0.0
ChPBV-S2-Nov1/2013-HUN	2	ORF1	RNA dependent RNA polymerase	Equine picobirnavirus Equ3	AKN50624	82%	100%	0.0
ChPBV-S2-Nov2/2013-HUN	2	ORF1	RNA dependent RNA polymerase	Equine picobirnavirus Equ1	AKN50618	57%	100%	0.0
ChPBV-S2-Nov3/2013-HUN	2	ORF1	RNA dependent RNA polymerase	Equine picobirnavirus Equ1	AKN50618	60%	100%	0.0
ChPBV-S2-Nov4/2013-HUN	2	ORF1	RNA dependent RNA polymerase	Dromedary picobirnavirus	AIY31287	86%	96%	0.0
ChPBV-S2-Nov5/2013-HUN	2	ORF1	RNA dependent RNA polymerase	Macaque picobirnavirus 25	AVD54061	72%	99%	0.0
ChPBV-S2-Nov6/2013-HUN	2	ORF1	RNA dependent RNA polymerase	Gorilla picobirnavirus	AVA30701	83%	100%	0.0

3.3. Phylogenetic analyses of ChPBV segments

The phylogenetic dendrograms based on the aa sequence alignments of capsid and RdRp of the study sequences and all of the known PBV sequences with the length of $>$ 450 aa (S1) and $>$ 500 aa (S2) show the clear separation of the S1 and S2 ChPBV sequences from each other and from the currently known other PBVs (Fig. 2). According to the phylogenetic positions of S2 study sequences the Nov4-Nov6 segments belong to the genogroup I (GI) while the other four S2 sequences (Nov1- 3 and ctg1042) belong to the GII group (Fig. 2). High-resolution pictures of the detailed phylogenetic trees are found in the Supplementary material (Fig. S2).
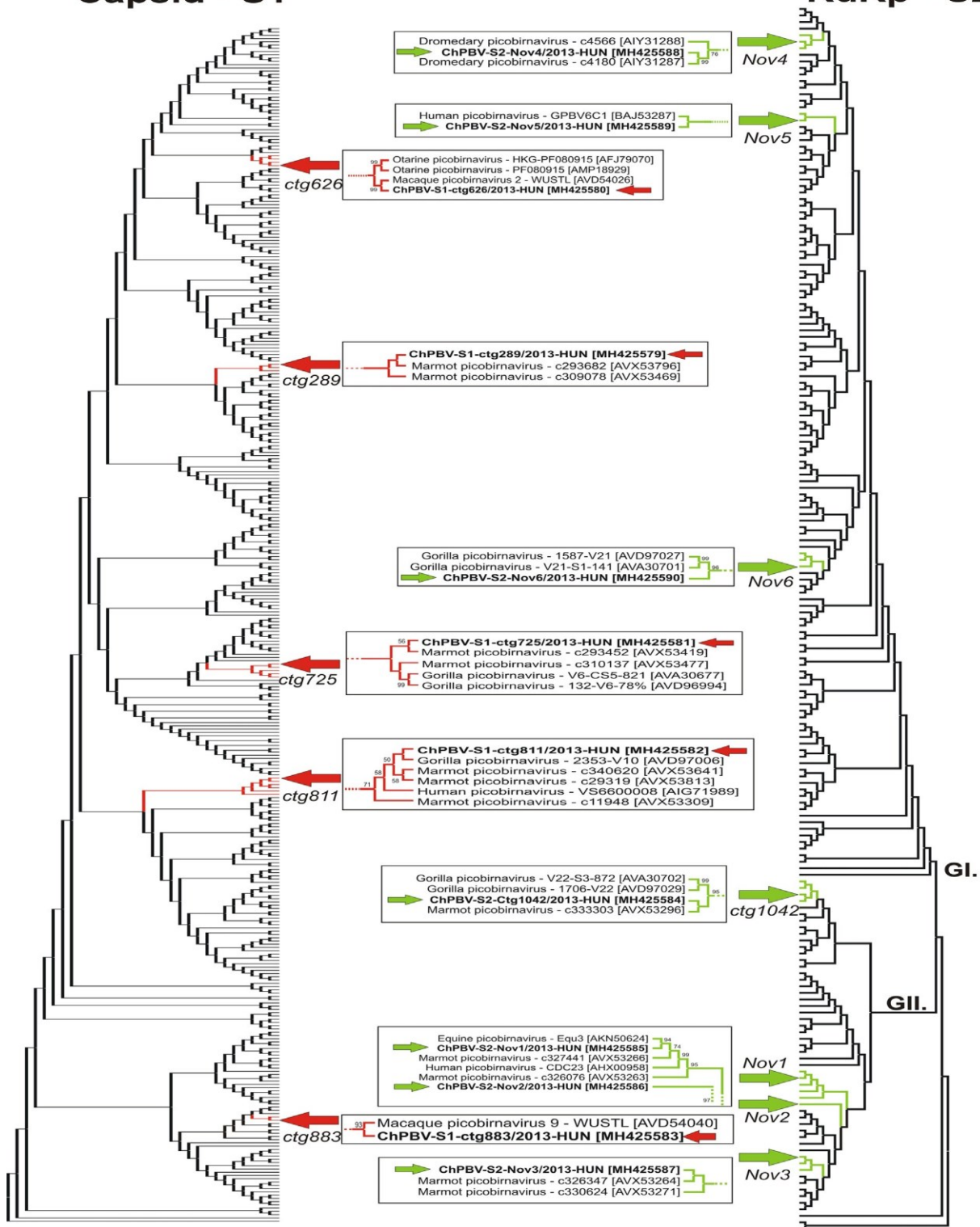
Fig. 2. The neighbor-joining phylogenetic trees based on the amino acid sequence alignments of capsid-S1 and RdRp-S2 of the study sequences (in bold) marked with red (S1 tree) and green arrows (S2 tree) and all of the known PBV sequences with the length of > 450 aa (capsid) and > 500 aa (RdRp). For clarity of the trees the sequence names of the complete trees are not shown. GI, GII: genogroups I and II. Branches which include the study sequences and closely related PBVs are depicted in detail including the accession numbers in brackets in framed pictures between the complete trees. High resolution version of the trees including all sequence names could be seen in Supplementary Fig. S2.

3.4. Putative pairing of ChPBV segments by RT-qPCR

Because there are different numbers of phylogenetically distant S1 and S2 segments identified from the analyzed cloacal sample pairing of the segments is challenging. Based on the highest sequence identity (9/7 nt, $\approx$ 78% identity), between the 3' terminals of S1 ctg811 and S2 ctg1042 (Table 2) these segments could belong to the same PBV and be packaged into the same virion. Because the S1 and S2 segments in the PBV virions are expected to be present in equimolar concentrations (Pereira et al., 1988; Green at al, 1999; Delmas, 2011) the pairing of S1- ctg811 and S2-ctg1042 could be supported by the similar Ct values of these segments measured by SYBR-green based RT-qPCR (Table 4). There are other S1 and S2 segments with similar Ct values (Table 4), although the nt comparison of the 3′ ends did not reveal any significant sequence identities between other theoretical segment pairs (Table 2, Table 4).

Table 4 Theoretical pairs of the identified segments (segments in same line) based on the similar Ct values (with $\pm$ standard deviations/SD/) measured by SYBR-Green based RT-qPCR. Note that the SD could be originated from the pipetting differences between the repeated measures.

Segment 1	Ct	Ct	Segment 2
ChPBV-S1-ctg811/2013-HUN	22.79 ± 0.89	22.09 ± 1.02	ChPBV-S2-Ctg1042/2013-HUN
		26.89 ± 0.75	ChPBV-S2-Nov1/2013-HUN
ChPBV-S1-ctg883/2013-HUN	28.23 ± 0.68	28.12 ± 0.67	ChPBV-S2-Nov2/2013-HUN
ChPBV-S1-ctg725/2013-HUN	29.85 ± 1.03	29.27 ± 0.91	ChPBV-S2-Nov3/2013-HUN
ChPBV-S1-ctg289/2013-HUN	24.44 ± 0.55	25.54 ± 0.79	ChPBV-S2-Nov4/2013-HUN
		27.18 ± 1.05	ChPBV-S2-Nov5/2013-HUN
ChPBV-S1-ctg626/2013-HUN	24.70 ± 1.05	25.63 ± 0.86	ChPBV-S2-Nov6/2013-HUN

3.5. Analysis of untranslated regions including presumed Shine-Dalgarno- like sequences of ChPBV segments

Analyses of nt alignments of the 5'UTRs of S1-ORF1 and S2-ORF1 sequences revealed the presence of the conserved 5' terminal G_1UAAA pentamer followed by an AU-rich region which is characteristic motifs of PBVs as well as other dsRNA viruses (Fig. 3) (Malik et al., 2014). Interestingly, one of the S2 sequences (Nov5) with the longest 5′UTR (75 nt, Table 2) contained a repeated sequence-region as an insert which could explain its unusually long 5'UTR (Fig. 3, Table 2). The presence of this 5′UTR stretch including the repeated region was verified by separate RT-PCR reaction and sequencing.

While the 5′ terminal UTRs of ChPBVs showed high variability in sequence lengths and identities a highly conserved heptamer (AGGAGG) $\approx$ 10–15 nt upstream to the initiation codon of the first ORFs of S1 and S2 was detected in all of the studied ChPBV sequences (Fig. 3). This conserved heptamer showed complete identity to the six-base consensus sequences (AGGAGG) of ribosomal-binding motifs (RBM) of prokaryotic mRNA molecules called Shine-Dalgarno (SD) sequences (Shine and Dalgarno, 1975) and similar to the RBMs of certain bacteriophages with segmented dsRNA genomes (cystoviruses) (Mindich, 1988) (Fig. 3). The SD-like sequences are also observable upstream of the initiation codons of the second and third ORFs of S1 sequences as recently reported in other PBVs (Krishnamurthy and Wang, 2018), although, the level of sequence conservation was de- creased compared to the SD-like sequences of S1-ORF1 or S2-ORF1 (Fig. 3). The predominant presence of SD-like sequences upstream of the presumed initiation codons suggests that the identified ChPBVs could be bacteriophages.
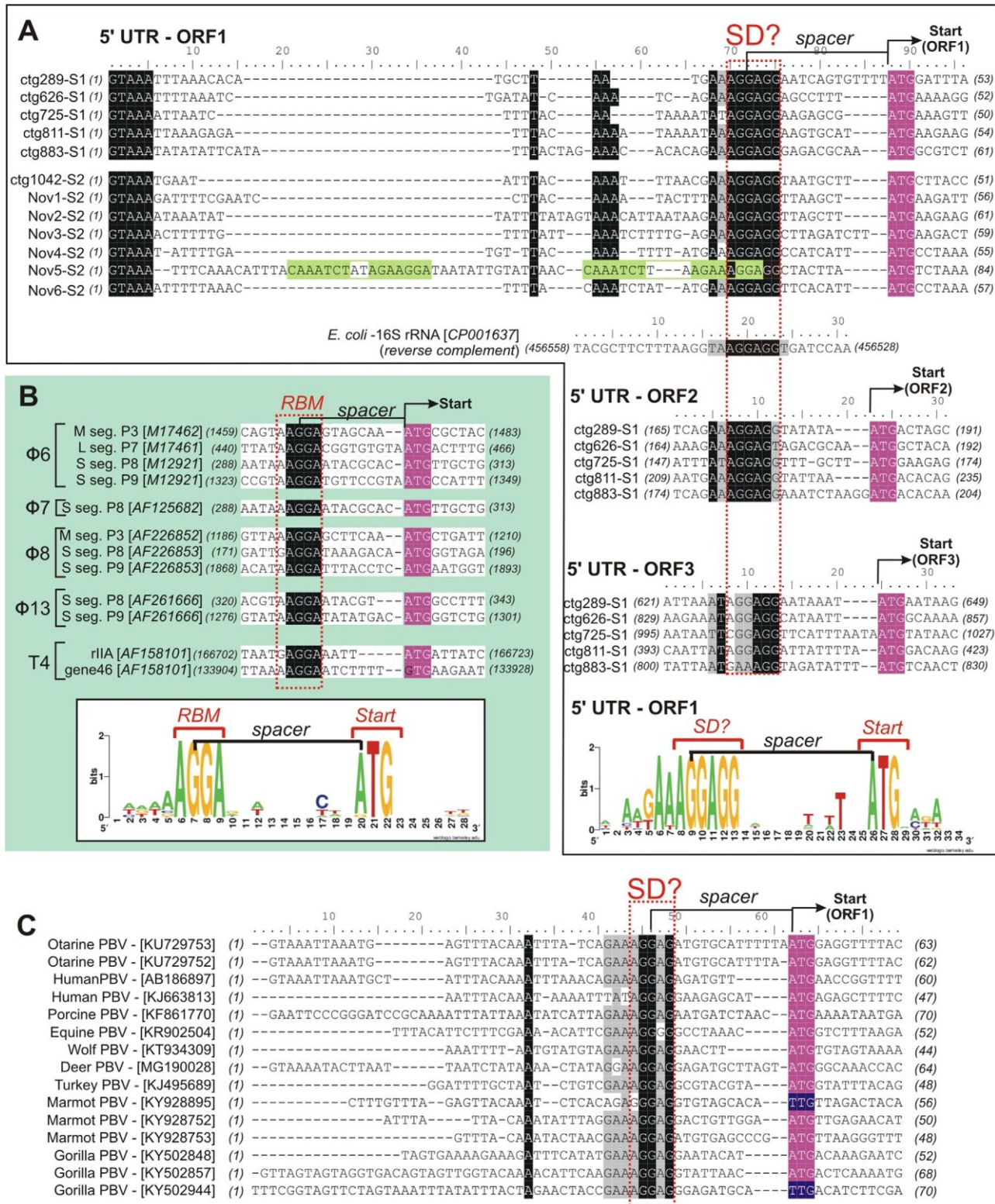
Fig. 3. (A) Nucleotide (nt) alignment of the 5′UTR sequences of the study segments and the Shine-Dalgarno (-like) sequences (SD?) located 10–15 nt upstream of every presumed ORFs and in 16S rRNA gene of *E. coli*. Nucleotides of a repeated sequence-region found in the 5′ UTR of S2 of Nov5 are framed and marked with green background. (B) Sequence alignment of genomic regions surrounded the ribosome-binding motifs (RBM) of representative dsRNA (Φ) and dsDNA phages (T4). Sequence logos of nt alignments of 5′ UTR-ORF1 sequences of the study segments and the Φ phages created by Weblogo (Crooks et al., 2004) are found at the bottom of the panels A and B. (C) nt alignment of the 5'UTR genomic regions including the SD-like sequences (SD?) located 8–15 nt upstream of the presumed ORF1 of S1 segments identified from different hosts.

Identical, highly conserved nts and presumed start codons are marked with black, grey and purple backgrounds in all panels. Note that alternative start codons (TTG) on panel C are marked with purple background.

3.6. Cultivation attempts of ChPBVs

To test the phage-hypothesis 3% solution of the original cloacal sample collected in 2013 and stored in -20 °C was inoculated to brain hearth infusion (BHI) broth and cultivated in aerobic and anaerobic conditions for up to two weeks to culture the attendant bacteria and ChPBVs potentially infecting the growing bacteria. The potential virus growth was monitored in regularly collected broth-samples using SYBR green-based RT-qPCR with the same segment-specific primers used for the segment-pairing reactions. None of the analyzed BHI samples collected from the two-week period showed any signs of ChPBV amplification detectable by qPCR (data not shown).

3.7. Functional analysis of SD-like sequences by His-tagged dsDNA construct and Western blot

To test the active nature of the detected SD-like sequences a recombinant dsDNA construct was generated by sequential PCR and ligation reactions using one of the identified ChPBV S1 sequences (ChPBV-S1-Ctg289/2013-HUN) as template and different oligonucleotide primers with 5′ 6x Histidine (6xHIS)-encoding tag (Fig. 4). For practical reasons beside the complete S1-ORF1 and S1-ORF2 only the 5′ 918 nt part of the S1-ORF3 was used in the construct (Fig. 4). The 1676- nt-long recombinant dsDNA construct contains a T7-promoter sequence at the 5′ end and the naturally presented, unaltered SD-like translation initiation sequences upstream of the ORFs and three C-terminal His- tagged protein-encoding viral ORFs (Fig. 4). The aim of this experiment was to transform a conventional cloning plasmid (pJET1.2blunt vector) into an *E. coli* expression system by use of the recombinant dsDNA construct as an insert (Fig. 4). If the viral SD-like sequences of the construct functioned as ribosomal binding site His-tagged viral proteins with known sizes (Fig. 4) should be expressed and became detectable in the lysate of the recombinant *E. coli* strains using Western blot (WB). Therefore the recombinant dsDNA (ctg289+) construct-containing pJET1.2blunt vector was used for the transformation of *E. coli* BL21(DE3)pLysS strain. The integrity of the insert was verified by sequencing in both directions of the plasmids isolated from the trans- formed *E. coli* cells. Recombinant protein expression was induced with 0.1 mM IPTG for 17 h. The soluble and pellet fractions of the transfected (ctg289+) and control (Neg.K) *E. coli* lysates were separated with 15% SDS-PAGE and the expressed peptides were detected with WB using 6x- His Tag-specific monoclonal antibody (Fig. 4). Although the immunoreactive bands corresponding to both the 16.77 kDa and 35.79 kDa viral peptides transcribed from the S1-ORF2 and S1-ORF3 were easily detectable after few seconds of exposure in the soluble and pellet fractions of the of the ctg289+ *E. coli* lysates (Fig. 4), the band corresponding the smallest (6.73 kDa) recombinant viral peptide transcribed from the presumed S1-ORF1 become visible only after an exposure time of 3 min in the pellet fraction of ctg289+ constructs indicating weak expression of this peptide (Fig. 4).
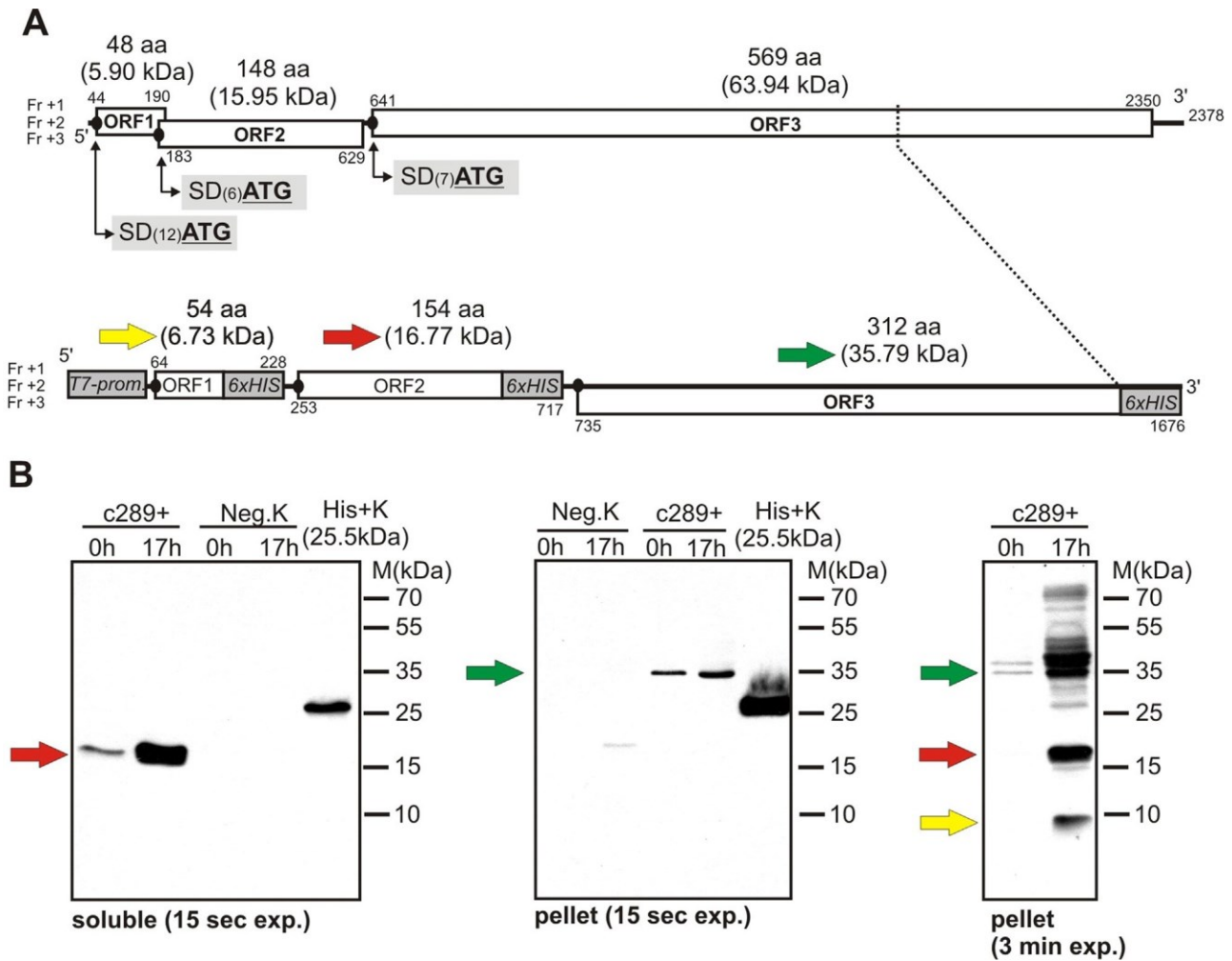
Fig. 4. (A) Schematic genome maps of intact S1 segment of ChPBV/ctg289/2013-HUN (above) and it's double stranded DNA (dsDNA) construct (below). The dsDNA construct which was used in the *E. coli* transformation experiments as an insert contains a 5' T7-promoter (T7 prom.) and N = 6 Histidine-codon (6xCAT) followed by stop codons (6xHIS) at the 3' ends of all three ORFs (ORF1-3). Due to practical reasons only the 5' 918 nt (306 aa) of the ORF3 was used in the synthetic dsDNA construct. Numbers above and below the open reading frame cassettes (ORF1-3) represents the first and last nucleotide (nt) of the ORF. The amino acid (aa) lengths and the calculated molecular masses (in brackets) of the peptides transcribed from the three ORFs are found above the given ORFs. SD: presumed Shine-Dalgarno-like sequences (black dots prior to ORFs) <u>ATG</u>: start codon of the given ORF. Numbers in brackets between SD and <u>ATG</u> show the lengths of the spacers. Fr +1, +2, +3: 1st, 2nd, 3rd reading frames. Yellow, Red and Green arrows indicate the molecular masses (on panel A) and the corresponding bands (on panel B) of the expressed peptides. (B): Expression and Western blot analysis of 6xHis-tagged recombinant ChPBV/ctg289/2013-HUN dsDNA construct (ctg289+) in *E. coli* BL21(DE3)pLysS strain. Beside the protein ladder (M), a 6xHis-tagged recombinant peptide with known molecular mass (25.5 kDa) was used for peptide-size prediction and for positive control of anti-His immunoreactivity (His+K). Immunoreactive bands marked with red, green and yellow arrows indicate the corresponding 16.77 kDa, 35.79 kDa and 6.73 kDa recombinant His-tagged viral peptides.

4. Discussion

In this study the complete sequences of twelve PBV segments were determined using next generation

sequencing and RT-PCR methods from a single cloacal sample of a diarrheic chicken held in a "back yard" family farm with inadequate hygienic conditions, in Hungary. Further background information about the sample and the farm see Boros et al., 2016 and: http://www.elsevierblogs.com/virology/unexpected- picornavirus-diversity-in-a-single-diarrheic-chicken/. According to our currently knowledge this is the highest number of completely determined PBV segments identified from a single specimen. Interestingly, the analyzed cloacal sample collected in 2013 also contained up to eight different chicken picornaviruses (family *Picornaviridae*) as well as other viruses belong to at least further six virus families indicating high level of enteric virus diversity in the case of chicken diarrhea (Boros et al., 2016). Unfortunately none of the identified PBV segments were detectable by SS-PAGE suggesting that the dsRNA concentration of the tested sample was below the detection limit of the applied SS-PAGE, which is plausible given the fact that PAGE is poorly ($\approx$ 100 fold less) sensitive than RT-PCR and detects only dsRNA viruses present in high viral load (Ganesh et al., 2014).

Based on the results of BLAST analyses as well as the presence of characteristic features of PBVs e.g. ExxRxNxxxE and capsid motifs of S1 and RdRp motifs of S2 (Da Costa et al., 2011; Duquerroy et al., 2009; Collier et al., 2016) from the identified twelve PBV sequences five are segment 1 and seven are segment 2 sequences.

The S1 and S2 sequences are phylogenetically distant from each other and from other currently known PBVs. While the identified S2 sequences belong to either GI or GII genogroups results of sequence analyses indicated that the widely used screening oligonucleotide primer sets for GI and GII PBVs (Carruyo et al., 2008; Rosen et al., 2000; Ganesh et al., 2012) would not be able to amplify these S2 study sequences.

The segment lengths of the identified segments show wide range mainly due to the sequence insertions/deletions (e.g. insertion of a repeated sequence-region in the 5'UTR of Nov5) and repeats (e.g. different number of ExxRxNxxxE motifs) similar as found in other PBVs (Da Costa et al., 2011; Bányai et al., 2014).

Although there are odd numbers of S1 (N = 5) and S2 (N = 7) segments identifiable in the analyzed sample some putative pairing of S1 and S2 segments are proposed based on similar Ct values measured by SYBR-Green based RT-qPCR. Furthermore in at least one case (S1- ctg811 and S2-ctg1042) the similar 3' terminal sequences could also support the proposed pairing (Table 2, Table 4). Conserved 3' ends which could be important for segment selection for virion packaging were described in other eukaryotic dsRNA viruses of family *Reoviridae* as well as dsRNA phages of family *Cystoviridae* (Attoui et al., 2012; Mindich, 1988). Conserved 3' ends of segments were also described in certain human, porcine and bovine PBVs (Malik et al., 2014; Bányai et al., 2014). However - based on the sequence variability of PBVs (Malik et al., 2014) - these theoretical segment pairs should be treated with caution

because of the potential presence of other PBV segments in the sample not identifiable by our initial BLASTx analyses. Furthermore, since reassortments among certain viruses with segmented genomes is common (Greenbaum et al., 2012), alternate segment sharing between different PBVs including ChPBVs may be possible (Bányai et al., 2014). Beside the well-known conserved terminal pentamers (G_1UAAA) (Malik et al., 2014) sequence analysis of the 5'UTRs showed the presence of conserved hexamers (AGGAGG) upstream of all of the presumed initiation codons which is identical to the ribosomal-binding motifs (RBM) of prokaryotic called Shine-Dalgarno sequences (Shine and Dalgarno, 1975). Similar RBMs are present in certain bacteriophages with segmented dsRNA genomes of family *Cystoviridae* (Mindich, 1988). Although the consistent presence of SD-like sequences of PBVs upstream of the S1-ORF2 and S1-ORF3 were recently reported (Krishnamurthy and Wang, 2018), during the preparation of this manuscript our sequence analyses reveal the presence of same SD-like sequences present upstream of the putative S1-ORF1 sequences of ChPBVs of this study as well as in other S1 segments of different PBVs (Fig. 3). Beside the sequence identities of SD-like sequences, the lengths (10–15 nt) of the spacers (sequences between the first nucleotides of the start codon and median guanosine of aag<u>G</u>agg) are in the same range as found in prokaryotic mRNAs as well as in the translation initiation sites of bacteria-infecting cystoviruses (Osterman et al., 2012; Mindich, 1988) (Fig. 3). These observations suggest that the identified ChPBVs could be bacteriophages.

In the present study cultivation attempts of ChPBVs using 1% solution of the cloacal sample inoculated into BHI broth were unsuccessful. The inoculum originated from 2013 and was initially taken in order to analyze viral nucleic acids rather than to maintain bacterial diversity or infectivity of viruses (i.e. not collected in a way to keep anaerobic conditions, and stored deep-frozen without preservatives). The number of cultivable, intact bacteria as well as the low initial infective particles may have therefore been significantly decreased.

In order to gather evidences supporting that picobirnaviruses are phages we tested the ability of the SD-like sequences present upstream of all presumed, *in silico* identified ORFs to initiate translation. For this propose a recombinant dsDNA construct from one of the identified S1 sequence (ChPBV-S1-Ctg289/2013-HUN) was created where 3' ends of all three S1-ORFs were modified to encode a 6xHis tag. Using this dsDNA construct a cloning plasmid was transformed into an expression plasmid for *E. coli* expression. After IPTG induction the expression of all three C-terminal 6xHis-tagged recombinant viral peptides encoded by the three S1-ORFs was detectable with Western blot using 6x-His Tag- specific monoclonal antibody. These results might indicate the in vivo functionality of the detected SD-like sequences in a bacterial *E. coli* system suggesting the presence of three functional ORFs in the S1 segments of ChPBVs. Because a presumed S1-ORF1 is also *in silico* identifiable in other PBVs (Fig. 3) the presence of three functional ORFs for those PBVs where an SD-like sequence was present upstream of the S1-ORF1 is plausible.

The results of this study together with the recently published data of other research groups (Krishnamurthy and Wang, 2018; Yinda et al., 2018) provide indirect evidences suggesting prokaryotic hosts for PBVs. Supporting possible eukaryotic cellular hosts for PBVs are mammalian viremia (Li et al., 2015), respiratory infections (Smits et al., 2011), liposome-perforating capacity of the virion and autoproteolytic processing of the PBV capsid (Duquerroy et al., 2009). Note those autoproteolytic capacities of bacteriophages (Wang et al., 2006) as well as detection of phages in respiratory samples (Lee and Bent, 2014) and other sterile environments such as blood (Dabrowska et al., 2005) have been described.

Our results should be interpreted in the light of some potential limitations (i) there are some mainly ancestral type mitochondria (e.g. protozoons like *Reclinomonas americana*) where the SD-like sequences were also abundant, and therefore could be the replication organelle of certain PBVs, although the leaderless metazoan mitochondrial mRNAs are generally not contains SD-sequences (Lang et al., 1997; Christian and Spremulli, 2010). (ii) Based on the presence of non-canonical translation-initiation mechanism in *E. coli* (e.g. leaky translation; Beck and Janssen, 2017) the expression of His-tagged recombinant peptides could be initiated by currently unknown SD-independent mechanisms. Therefore, until successful cultivation of PBVs in specific bacterial cultures is achieved the phage nature of PBVs remains hypothetical. Given the fact that gut microbiome consist of several hundreds of mostly uncultivable bacteria (Kim et al., 2011) the identification of true bacterial or archeal host(s) of PBVs (if any) will be challenging. Based on the present update of knowledge searches for potential host of PBVs should be extended to prokaryotes in the bacterial gut flora.

Table S1: Pairwise identity values of nucleotide (nt) and amino acid (aa) sequences of the study ChPBV segments. The highest identity values are marked with bold.

Segment 1

	Ctg289_ORF1	Ctg626_ORF1	Ctg725_ORF1	Ctg811_ORF1	Ctg883_ORF1	nt
Ctg289_ORF1		46.93%	46.93%	42.85%	35.46%	
Ctg626_ORF1	27.08%		37.77%	34.52%	**47.51%**	
Ctg725_ORF1	18.75%	22.03%		41.66%	41.84%	
Ctg811_ORF1	10.41%	9.09%	9.09%		35.46%	
Ctg883_ORF1	13.04%	13.04%	8.69%	13.04%		
aa	Ctg289_ORF1	Ctg626_ORF1	Ctg725_ORF1	Ctg811_ORF1	Ctg883_ORF1	

Segment 1

	Ctg289_ORF2	Ctg626-ORF2	Ctg725_ORF2	Ctg811_ORF2	Ctg883_ORF2	nt
Ctg289_ORF2		37.23%	30.60%	30.81%	37.10%	
Ctg626-ORF2	15.89%		**42.06%**	25.86%	38.16%	
Ctg725_ORF2	16.00%	**23.48%**		18.37%	35.86%	
Ctg811_ORF2	20.95%	12.90%	10.61%		24.45%	
Ctg883_ORF2	15.16%	22.33%	20.76%	15.56%		
aa	Ctg289_ORF2	Ctg626-ORF2	Ctg725_ORF2	Ctg811_ORF2	Ctg883_ORF2	

Segment 1

	Ctg298_ORF3	Ctg626-ORF3	Ctg725-ORF3	Ctg811_ORF3	Ctg883_ORF3	nt
Ctg298_ORF3		42.52%	41.72%	41.66%	40.12%	
Ctg626-ORF3	25.02%		41.00%	42.22%	38.92%	
Ctg725-ORF3	24.51%	**25.82%**		**45.66%**	38.45%	
Ctg811_ORF3	24.67%	22.66%	24.76%		37.90%	
Ctg883_ORF3	23.46%	19.10%	19.74%	18.89%		
aa	Ctg298_ORF3	Ctg626-ORF3	Ctg725-ORF3	Ctg811_ORF3	Ctg883_ORF3	

Segment 2

|	Ctg1042_ORF	Nov1_ORF	Nov2_ORF	Nov3_ORF	Nov4_ORF	Nov5_ORF	Nov6_ORF	nt
Ctg1042_ORF		42.78%	44.52%	43.12%	35.58%	35.44%	37.94%	
Nov1_ORF	29.63%		49.49%	49.21%	35.92%	34.12%	37.36%	
Nov2_ORF	30.04%	47.83%		55.40%	37.06%	36.58%	39.08%	
Nov3_ORF	30.29%	49.47%	56.29%		37.15%	35.54%	38.26%	
Nov4_ORF	19.57%	20.91%	22.15%	21.87%		55.36%	**60.04%**	
Nov5_ORF	19.04%	21.77%	21.23%	21.54%	58.67%		56.93%	
Nov6_ORF	20.79%	22.85%	21.73%	22.77%	61.68%	**62.25%**		
aa	Ctg1042_ORF	Nov1_ORF	Nov2_ORF	Nov3_ORF	Nov4_ORF	Nov5_ORF	Nov6_ORF	
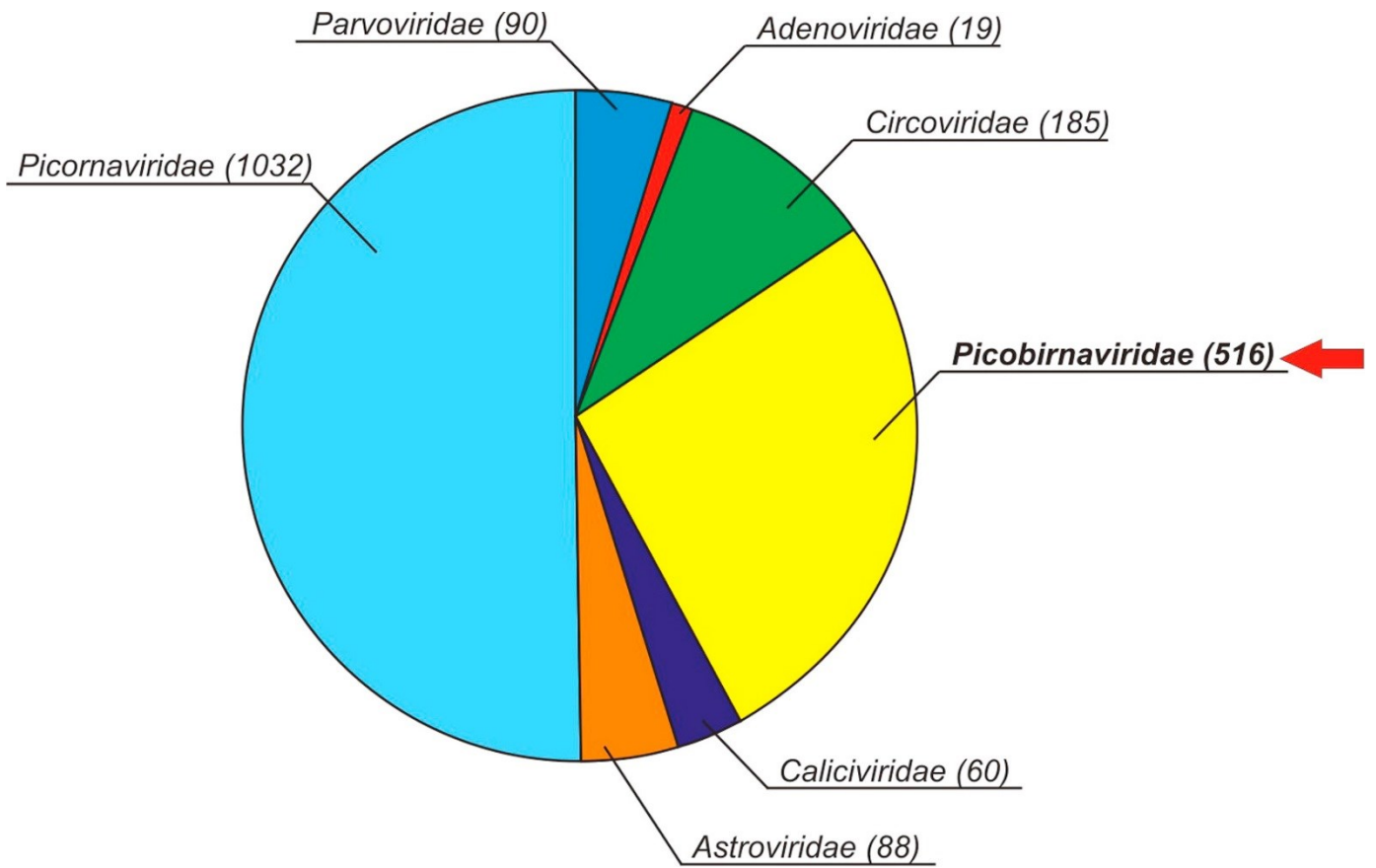
Fig. S1. Pie-chart illustrating the relative abundance of viral metagenomic reads belong to different eukaryotic virus families. Virus family-level categorization was based on the results of BLASTx analyses. Numbers in brackets represent the number of reads. The picobirnavirus-related reads are in bold and marked with an arrow.
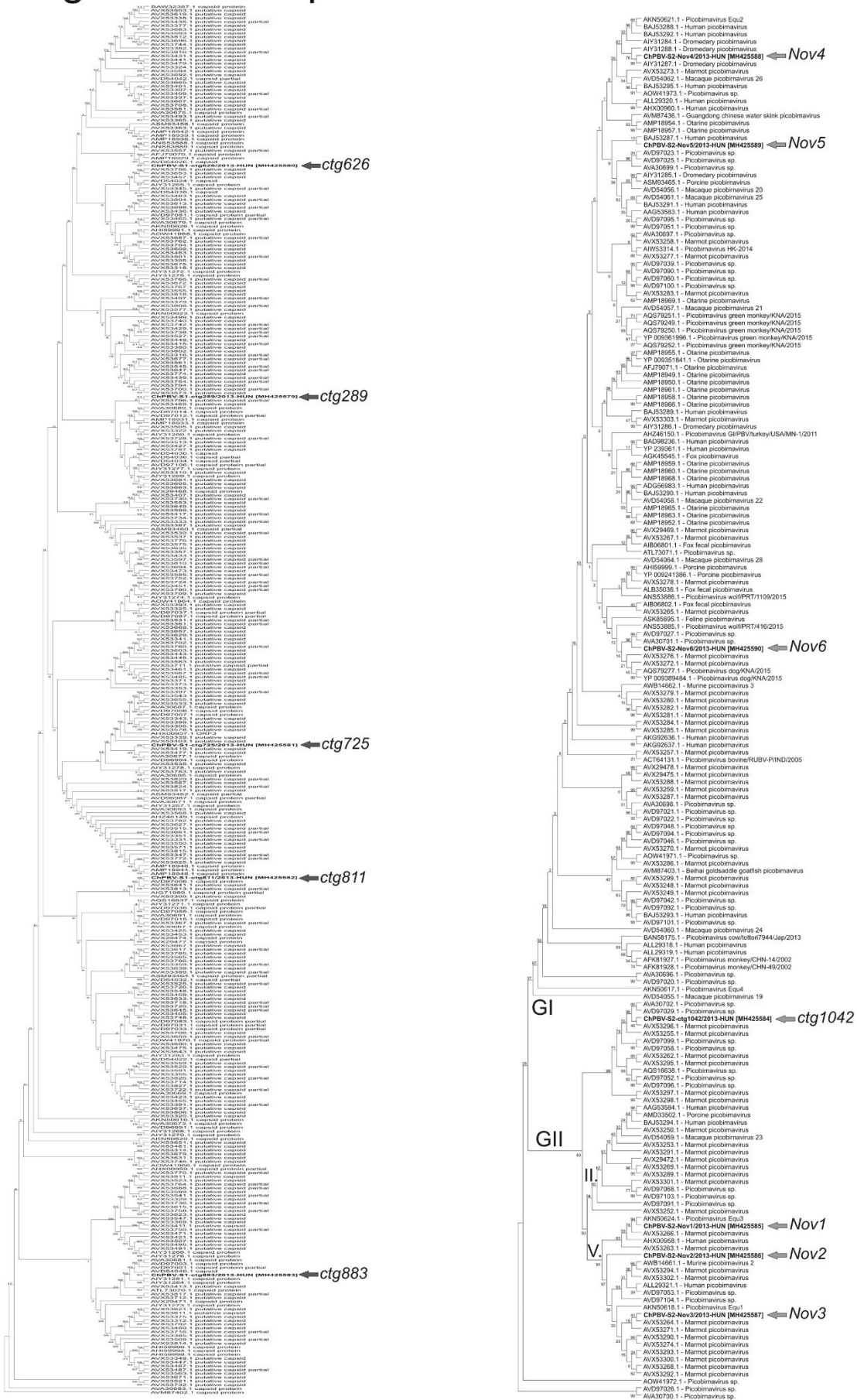
Segment 1 - Capsid

Segment 2 - RdRp

Fig. S2. High resolution version of the neighbor-joining phylogenetic trees including all sequence names based on

the amino acid sequence alignments of Segment1-capsid and Segment2-RdRp of the study sequences (in bold) and all of the known PBV sequences with the length of > 450 aa (capsid) and > 500 aa (RdRp). GI, GII: genogroups I and II.

References

Anthony, S.J., Islam, A., Johnson, C., Navarrete-Macias, I., Liang, E., Jain, K., Ojeda- Flores, R., 2015. Non-random patterns in viral diversity. Nat. Commun. 6, 8147.

Attoui, H., Mertens, P.P.C., Becnel, J., Belaganahalli, S., Bergoin, M., Brussaard, C.P., Zhou, H., 2012. Part II: the viruses – the double stranded RNA viruses - family *Reoviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, pp. 541–637.

Bányai, K., Potgieter, C., Gellért, Á., Ganesh, B., Tempesta, M., Lorusso, E., Martella, V., 2014. Genome sequencing identifies genetic and antigenic divergence of porcine picobirnaviruses. J. Gen. Virol. 95 (10), 2233–2239.

Beck, H.J., Janssen, G.R., 2017. Novel translation initiation regulation mechanism in Escherichia coli ptrB mediated by a 5′-terminal AUG. J. Bacteriol. JB 00091.

Boros, Á., Pankovics, P., Adonyi, Á., Fenyvesi, H., Day, J.M., Phan, T.G., Reuter, G., 2016. A diarrheic chicken simultaneously co-infected with multiple picornaviruses: complete genome analysis of avian picornaviruses representing up to six genera. Virology 489, 63–74.

Carruyo, G.M., Mateu, G., Martínez, L.C., Pujol, F.H., Nates, S.V., Liprandi, F., Ludert, J.E., 2008. Molecular characterization of porcine picobirnaviruses and development of a specific reverse transcription-PCR assay. J. Clin. Microbiol. 46 (7), 2402–2405.

Christian, B.E., Spremulli, L.L., 2010. Preferential selection of the 5'terminal start codon on leaderless mRNAs by mammalian mitochondrial Ribosomes. J. Biol. Chem. jbc- M110.

Collier, A.M., Lyytinen, O.L., Guo, Y.R., Toh, Y., Poranen, M.M., Tao, Y.J., 2016. Initiation of RNA polymerization and polymerase encapsidation by a small dsRNA virus. PLoS Pathog. 12 (4), e1005523.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14 (6), 1188–1190.

Da Costa, B., Duquerroy, S., Tarus, B., Delmas, B., 2011. Picobirnaviruses encode a protein with repeats of the ExxRxNxxxE motif. Virus Res. 158 (1–2), 251–256.

Dabrowska, K., Switała-Jelen, K., Opolski, A., Weber-Dabrowska, B., Gorski, A., 2005. Bacteriophage penetration in vertebrates. J. Appl. Microbiol. 98 (1), 7–13.

Day, J.M., Zsak, L., 2013. Recent progress in the characterization of avian enteric viruses. Avian Dis. 57 (3), 573–580.

Day, J.M., Oakley, B.B., Seal, B.S., Zsak, L., 2015. Comparative analysis of the intestinal bacterial and RNA viral communities from sentinel birds placed on selected broiler chicken farms. PLoS One 10 (1), e0117210.

Delmas, B., 2011. Family *Picobirnaviridae*. In: Proceedings of the Ninth Report of the International Committee on Taxonomy of Viruses, pp. 535–539.

Duquerroy, S., Da Costa, B., Henry, C., Vigouroux, A., Libersou, S., Lepault, J., Rey, F.A., 2009. The picobirnavirus crystal structure provides functional insights into virion assembly and cell entry. EMBO J. 28 (11), 1655–1665.

Duraisamy, R., Akiana, J., Davoust, B., Mediannikov, O., Michelle, C., Robert, C., Desnues, C., 2018. Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. Virus Genes 1–16.

Ganesh, B., Bányai, K., Martella, V., Jakab, F., Masachessi, G., Kobayashi, N., 2012. Picobirnavirus infections: viral persistence and zoonotic potential. Rev. Med. Virol. 22 (4), 245–256.

Ganesh, B., Masachessi, G., Mladenova, Z., 2014. Animal picobirnavirus. Virus Dis. 25 (2), 223–238.

Green, J., Gallimore, C.I., Clewley, J.P., Brown, D.W.G., 1999. Genomic characterisation of the large segment of a rabbit picobirnavirus and comparison with the atypical picobirnavirus of Cryptosporidium parvum. Arch. Virol. 144 (12), 2457–2465.

Greenbaum, B.D., Li, O.T., Poon, L.L., Levine, A.J., Rabadan, R., 2012. Viral reassortment as an information exchange between viral segments. Proc. Natl. Acad. Sci. 109 (9), 3341–3346.

Herring, A.J., Inglis, N.F., Ojeh, C.K., Snodgrass, D.A., Menzies, J.D., 1982. Rapid diagnosis of rotavirus infection by direct detection of viral nucleic acid in silver-stained polyacrylamide gels. J. Clin. Microbiol. 16 (3), 473–477.

Kim, B.S., Kim, J.N., Cerniglia, C.E., 2011. In vitro culture conditions for maintaining a complex population of human gastrointestinal tract microbiota. BioMed. Res. Int. 2011.

Krishnamurthy, S.R., Wang, D., 2018. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. Virology 516, 108–114.

Lang, B.F., Burger, G., O'kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Gray, M.W., 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387 (6632), 493–497.

Lee, C.K., Bent, S.J., 2014. Uncovering the hidden villain within the human respiratory microbiome. Diagnosis 1 (3), 203–212.

Li, L., Giannitti, F., Low, J., Keyes, C., Ullmann, L.S., Deng, X., Delwart, E., 2015. Exploring the virome of diseased horses. J. Gen. Virol. 96 (9), 2721–2733.

Luo, X.L., Lu, S., Jin, D., Yang, J., Wu, S.S., Xu, J., 2018. Marmota himalayana in the Qinghai–Tibetan plateau as a special host for bi-segmented and unsegmented pico-birnaviruses. Emerg. Microbes Infect. 7 (1), 20.

Malik, Y.S., Kumar, N., Sharma, K., Dhama, K., Shabbir, M.Z., Ganesh, B., Banyai, K., 2014.

Epidemiology, phylogeny, and evolution of emerging enteric Picobirnaviruses of animal origin and their relationship to human strains. BioMed. Res. Int. 2014.

Mindich, L., 1988. Bacteriophage Ø6: a unique virus having a lipid-containing membrane and a genome composed of three dsRNA segments. In: Maramorosch, K., Murphy, F.A., Shatkin, A.J. (Eds.), Advances in Virus Research Vol. 35. Academic Press, San Diego, CA, pp. 137–173 (ISBN: 0-12-039835-4).

Ng, T.F.F., Vega, E., Kondov, N.O., Markey, C., Deng, X., Gregoricus, N., Delwart, E., 2014. Divergent picobirnaviruses in human feces. Genome Announc. 2 (3), e00415–14.

Osterman, I.A., Evfratov, S.A., Sergiev, P.V., Dontsova, O.A., 2012. Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res. 41 (1), 474–486.

Pereira, H.G., Fialho, A.M., Flewett, T.H., Teixeira, J.M.S., Andrade, Z.P., 1988. Novel viruses in human faeces. Lancet 332 (8602), 103–104.

Rosen, B.I., Fang, Z.Y., Glass, R.I., Monroe, S.S., 2000. Cloning of human picobirnavirus genomic segments and development of an RT-PCR detection assay. Virology 277 (2), 316-329.

Shine, J., Dalgarno, L., 1975. Determinant of cistron specificity in bacterial ribosomes. Nature 254 (5495), 34.

Silva, R.R., Bezerra, D.A.M., Kaiano, J.H.L., de Souza Oliveira, D., Silvestre, R.V.D., Gabbay, Y.B., Mascarenhas, J.D.A.P., 2014. Genogroup I avian picobirnavirus detected in Brazilian broiler chickens: a molecular epidemiology study. J. Gen. Virol. 95 (1), 117–122.

Smits, S.L., Poon, L.L., van Leeuwen, M., Lau, P.N., Perera, H.K., Peiris, J.S.M., Osterhaus, A.D., 2011. Genogroup I and II picobirnaviruses in respiratory tracts of pigs. Emerg. Infect. Dis. 17 (12), 2328.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Wakuda, M., Pongsuwanna, Y., Taniguchi, K., 2005. Complete nucleotide sequences of two RNA segments of human picobirnavirus. J. Virol. Methods 126 (1–2), 165–169.

Wang, S., Chang, J.R., Dokland, T., 2006. Assembly of bacteriophage P2 and P4 pro- capsids with internal scaffolding protein. Virology 348 (1), 133–140.

Woo, P.C., Lau, S.K., Teng, J.L., Tsang, A.K., Joseph, M., Wong, E.Y., Wernery, U., 2014. Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. Virology 471, 117–125.

Yinda, C.K., Ghogomu, S.M., Conceição-Neto, N., Beller, L., Deboutte, W., Vanhulle, E., Matthijnssens, J., 2018. Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. Virus Evol. 4 (1), vey008.