Virology Journal

SHORT REPORT

Multiple novel filamentous phages detected in the cloacal swab samples of birds using viral metagenomics approach

Jian Zeng[†], Yan Wang[†], Ju Zhang, Shixing Yang and Wen Zhang[*]

Abstract

Members of the family *Inoviridae* (inoviruses) are characterized by their unique filamentous morphology and infection cycle. The viral genome of inovirus is able to integrate into the host genome and continuously releases virions without lysing the host, establishing chronic infection. A large number of inoviruses have been obtained from microbial genomes and metagenomes recently, but putative novel inoviruses remaining to be identified. Here, using viral metagenomics, we identified four novel inoviruses from cloacal swab samples of wild and breeding birds. The circular genome of those four inoviruses are 6732 to 7709 nt in length with 51.4% to 56.5% GC content and encodes 9 to 13 open reading frames, respectively. The zonula occludens toxin gene implicated in the virulence of pathogenic host bacteria were identified in all four inoviruses and shared the highest amino acid sequences identity (< 37.3%) to other reference strains belonging to different genera of the family *Inoviridae* and among themselves. Phylogenetic analysis indicated that all the four inoviruses were genetically far away from other strains belonging to the family *Inoviridae* and formed an independent clade. According to the genetic distance-based criteria, all the four inoviruses identified in the present study respectively belong to four novel putative genera in the family *Inoviridae*.

Keywords: Viral metagenomics, Cloacal swab, Filamentous phage, *Inoviridae*, Virus evolution

Main text

The *Inoviridae* is a large family of non-enveloped, flexible filamentous viruses with circular, +ssDNA genomes of about 5.5 to 10.6 kb in size which encodes 7–15 proteins. The family *Inoviridae* is divided into 21 genera including *Affertcholeramvirus, Capistrivirus, Coriovirus, Fibrovirus, Habenivirus, Infulavirus, Inovirus, Lineavirus, Parhipatevirus, Primolicivirus, Psecadovirus, Restivirus, Saetivirus, Scuticavirus, Staminivirus, Subteminivirus, Tertilicivirus, Versovirus, Vicialiavirus, Villovirus*, and *Xylivirus* (www. ictv.global/report/inoviridae). They infect gram-negative bacteria by using a unique strategy of virion morphogenesis [1–4]. They adsorb to host bacterial pill that are thought to spontaneously resulting in the entry of the phage genome into the host cytoplasm. Following entry, filamentous phage genomes replicate through a rolling-circle mechanism. Progeny virions release from cells by extrusion without causing host lysis. Filamentous phage DNA can persist extra-chromosomally or integrate into the bacterial genome. The interaction between filamentous phages and their hosts is generally considered to be a symbiotic relationship, phage replication only causes a slight burden on the host, and in turn increases the virulence and toxicity of the host evolution. Most of our understanding of the biology of filamentous phages comes from *Escherichia coli* phages called Ff filamentous phages which were discovered in sewage samples in the early 1960s [5]. Ff filamentous phage is one of the early cloning vectors for DNA sequencing, and served as the

*Correspondence: z0216wen@yahoo.com
[†]Jian Zeng and Yan Wang have contributed equally to this work
School of Medicine, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, Jiangsu, People's Republic of China

Zeng *et al. Virology Journal*   (2021) 18:240

Page 2 of 6

workhorse of phage display technology for past 30 years [6, 7]. Other filamentous phages infected *Pseudomonas aeruginosa* were called the "Pf1-like" phages (Pf1, Pf4, Pf5, Pf6 and Pf7) and widespread among *P. aeruginosa* strains [8]. Pf4 has a key role in the overall structure, organized remodeling and seeding of mature biofilm. Filamentous phages existing in biofilm matrix are self-organized into viscous liquid crystal arrangement, which provides bacteria with enhanced surface adhesion and resistance to drying and antibiotics [9]. In addition, Pf phage may contribute to clinical outcomes in *P. aeruginosa* infection in patients with cystic fibrosis [10]. A large number of inovirus-like sequences were obtained from microbial genomes and metagenomes recently, but still have many filamentous phages need to be identified [11, 12]. Wild birds as one of the most abundant species carry a large number of viruses. Here, four novel filamentous phages divergent with existing members of the family *Inoviridae* were identified from viral metagenomics database of wild bird.
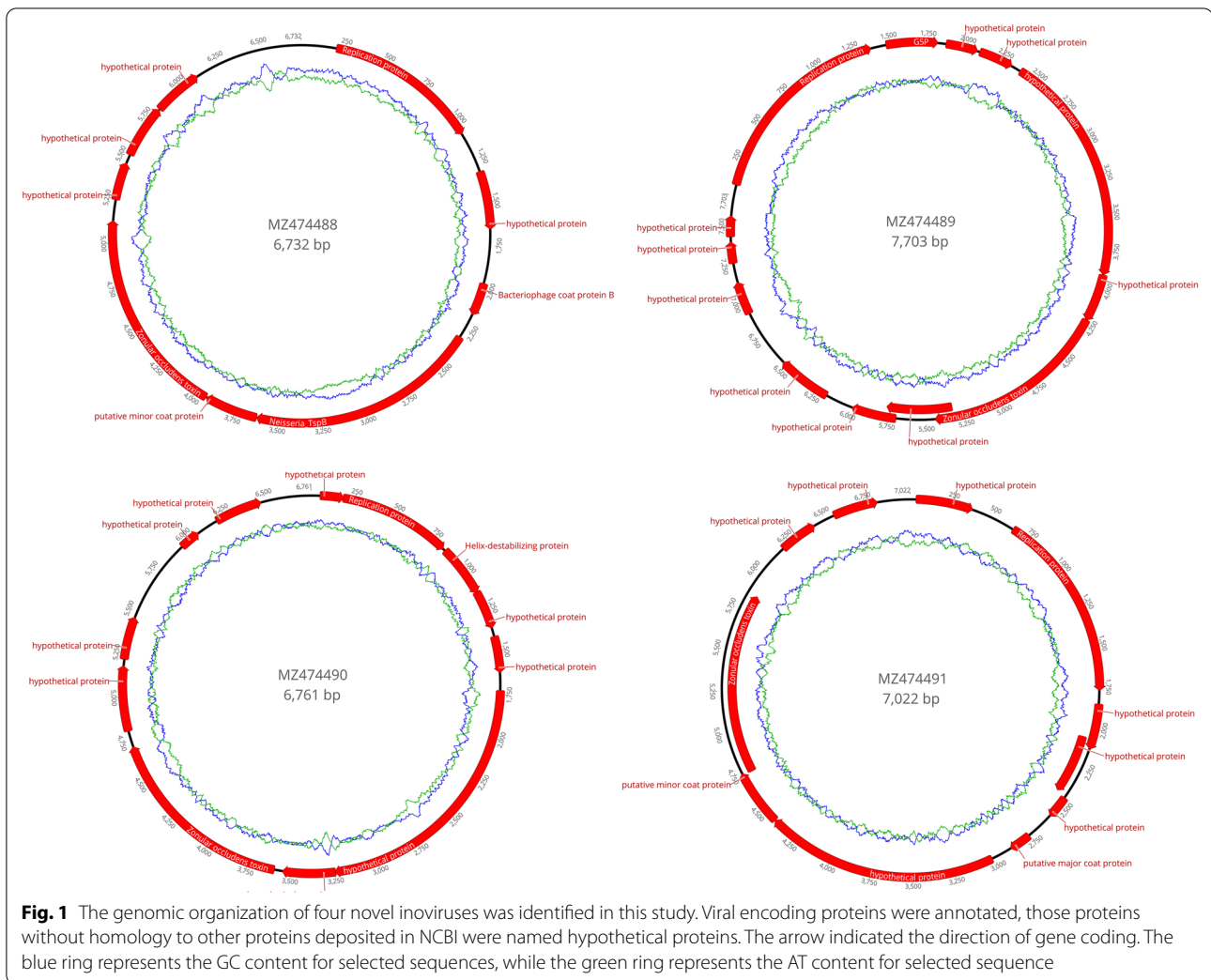
During 2018 to 2019, 43 cloacal swabs of wild and breeding bird samples were collected by using disposable absorbent cotton swabs from Hunan and Zhejiang province of China including 16 for Black Swans (Hunan, wild), 15 for Sliver Pheasant (Hunan, wild), and 16 for Toucan (Zhejiang, breeding). All of the birds were adults and the exact ages were unknown.

Viral metagenomics approach was used to characterize viral sequences in the fecal samples. Briefly, tips of swabs were immersed into 0.5 ml Dublecco's phosphate-buffered saline (DPBS) and vigorously vortexed for 10 min, then centrifuged at $15,000 \times g$ for 10 min. Each supernatant was collected in a new 1.5 ml centrifuge tubes and stored at -80℃ until use. About 50 μl supernatant from each sample was pipetted and pooled into different sample pools. Three sample pools were generated based on the bird species. Sample pools were centrifuged at $12,000 \times g$ for 20 min at 4℃ and then filtered through a 0.45-μm filter to remove eukaryotic and bacterial cell-sized particles. Filtrates were digested by DNase and RNase at 37℃ for 60 min. Total nucleic acids were then extracted using QIAamp MinElute Virus Spin Kit (Qiagen) according to the manufacturer's protocol. Three libraries were constructed using Nextera XT DNA Sample Preparation Kit (Illumina) and sequences using the Miseq Illumina platform with 250 bases paired ends with a distinct molecular tag for each pool. For bioinformatics analysis, resulting raw reads were debarcoded using vendor software from Illumina. An in-house analysis pipeline running on a 32-nodes Linux cluster was used to process the data. Reads were considered duplicates if bases 5 to 55 were identical and only one random copy was kept. Clonal reads were removed and low sequencing

quality tails were trimmed using Phred quality score ten as the threshold. Adaptors were trimmed using the default parameters of VecScreen which is NCBI BLASTn with specialized parameters designed for adaptor removal. Bacterial reads were subtracted by mapping to the bacterial nucleotide sequences from the BLAST NT database using Bowtie2 v2.2.4. The cleaned reads were de-novo assembled by SOAPdenovo2 version r240 using Kmer size 63 with default settings [13]. The assembled contigs, along with singlets were aligned to an in-house viral proteome database using BLASTx (v.2.2.7) with an E-value cutoff of $< 10^{-5}$, where the virus BLASTx database was compiled using NCBI virus reference proteome (https://ftp.ncbi.nih.gov/refseq/release/viral/) to which was added viral protein sequences from NCBI nr fasta file (based on annotation taxonomy in Virus Kingdom).

For phylogenetic analysis, the zonula occludens toxin (Zot) and major coat protein sequences of reference strains belonging to different genera of *Inoviridae* were downloaded from the NCBI GenBank database. Related protein sequences were aligned alignment program implemented in the CLC Genomics Workbench 10.0, and the resulting alignment was further optimized using MUSCLE in MEGA-X [14] and MAFFT v7.3.1 employing the E-INS-I algorithm [15]. Sites containing more than 50% gaps were temporarily removed from alignments. Bayesian inference trees were then constructed using MrBayes v3.2.7 [16]. The Markov chain was run for a maximum of one million generations, in which every 50 generations were sampled and the first 25% of Markov chain Monte Carlo samples were discarded as burn-in.

Results showed that 599, 1236, and 1474 sequence reads respectively from 3 libraries belonged to the family *Inoviridae*. Four complete genomes of inoviruses were obtained by assembling using the low sensitivity/fastest parameter in Geneious 11.1.2 and named black-swan219-1, blackswan219-6, silverpheasant213, and toucan80 separately. Those four novel inoviruses were submitted to GenBank with accession no. MZ474488 to MZ474491. The genome of those four inoviruses are 6732 nt, 6761 nt, 7022 nt, and 7709 nt in length and have 9 to 13 open reading frames (ORF) respectively (Fig. 1). The GC contents of those four inoviruses are 56.5% for MZ474489, 54.3% for MZ474491, 51.4% for MZ474490, and 56.4% for MZ474488. Most encoding proteins of those inoviruses have no homology with other proteins deposited in NCBI database, while the replication protein (Rep) and Zot similar to other phages were detected in all those four inoviruses. The Rep of those four inoviruses shared amino acid sequence identity (98.3%, 76.5%, 94.4%, and 56.5%) to different Pseudomonas sp. deposited in NCBI. The Zot of those four inoviruses shared amino acid sequence identity (89.7%, 92.4%, 63.6%, and

Zeng *et al. Virology Journal*   (2021) 18:240

Page 3 of 6

Fig. 1 The genomic organization of four novel inoviruses was identified in this study. Viral encoding proteins were annotated, those proteins without homology to other proteins deposited in NCBI were named hypothetical proteins. The arrow indicated the direction of gene coding. The blue ring represents the GC content for selected sequences, while the green ring represents the AT content for selected sequence

30.3%) to different Pseudomonas sp. respectively. In addition, the putative major coat protein (CoaB) was only found in MZ474488 and MZ474491. Based on the above results, the Pseudomonas sp. may be the natural host of those four novel inoviruses and still need further study to testify it. Recently, analysis of metagenomic sequences has become an important means to study the diversity of viruses including eukaryotic viruses, bacterial viruses and archaeal viruses [17, 18]. Eugene V. Koonin et al., identified numerous genomes of virus-like elements that similar in size to tectiviruses and have diverse gene composition through searching tectivirus-like Double Jelly-Roll major capsid proteins in genomic and metagenomics sequence databases [19]. The Zot protein which conservative in the family *Inoviridae* can uses as the target for searching inoviruses in submitted metagenomics sequence databases.

To determine the relationship between those four inoviruses and members of the family *Inoviridae*, the amino acid sequences of Zot and CoaB were compared and analyzed (Fig. 2). The result showed that the Zot proteins of all those four inoviruses shared the highest amino acid sequences identity (< 37.5%) to other reference strains belonging to different genera of the family *Inoviridae* and between each other (Fig. 2a), while the CoaB of MZ474488 and MZ474491 had the highest amino acid sequences identity of 45.2% and 70.5% to other reference strains (Fig. 2b). The ICTV states that inoviruses in different genera differ from each other by > 50% in the amino acid sequence of CoaB and Zot as assigned (ttps://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/inoviridae). Based on the criteria, those four inoviruses identified here belongs to four novel genera of *Inoviridae*. In recent years, a genome organization-based taxonomy of prokaryotic viruses substituted for traditional morphology-based classifications to provide a relatively robust classification guide even for viruses with highly divergent genome sequences and organizations.

Zeng *et al. Virology Journal*   (2021) 18:240

Page 4 of 6

a Percent Identity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		
1	■	37.5	11.1	12.6	5.7	12.5	13.0	7.1	10.9	12.0	11.1	8.8	11.8	22.9	11.3	30.7	8.6	13.0	21.3	7.6	9.5	28.8	1	MZ474489 ●
2	109.5	■	13.9	12.3	3.0	10.2	11.6	8.1	13.5	9.9	8.4	5.8	10.8	17.4	8.6	26.0	13.9	9.1	18.7	8.1	11.1	21.8	2	MZ474488 ●
3	352.0	365.0	■	11.8	4.2	12.2	10.5	20.0	12.9	15.1	19.2	22.4	16.5	13.8	25.3	11.8	21.0	15.1	10.9	4.7	16.1	13.5	3	NC_001331-Primolicivirus
4	451.0	360.0	380.0	■	5.2	58.1	42.7	10.4	52.0	12.9	13.2	13.2	9.3	13.7	5.8	7.4	11.5	5.8	11.8	4.1	17.5	7.9	4	NC_001332-Lineavirus
5	661.0	1000.0	1000.0	605.0	■	3.7	2.5	2.6	4.3	6.5	2.6	9.1	4.7	3.4	7.8	5.8	4.0	2.1	6.1	6.7	5.5	8.4	5	NC_001418-Tertilicivirus
6	493.0	426.0	403.0	61.8	560.0	■	40.8	11.0	65.8	12.7	12.5	9.9	6.5	9.9	5.9	12.2	11.0	7.9	9.6	4.0	15.6	6.2	6	NC_001954-Infulavirus
7	531.0	383.0	328.0	94.1	545.0	105.2	■	11.1	44.8	13.0	13.0	4.4	5.5	12.7	8.9	8.0	4.7	11.4	5.5	4.7	11.6	10.5	7	NC_001956-Saetivirus
8	335.0	332.0	236.0	393.0	1000.0	360.0	332.0	■	12.4	16.1	13.9	39.8	10.0	9.3	21.1	11.8	41.2	14.0	9.6	6.3	14.5	10.0	8	NC_002362-Versovirus
9	524.0	476.0	407.0	74.6	545.0	45.1	95.0	347.0	■	17.8	12.9	11.5	10.3	10.6	9.2	12.4	12.9	10.3	6.9	3.2	17.0	11.8	9	NC_003287-Inovirus
10	403.0	456.0	347.0	945.0	349.0	328.0	393.0	389.0	349.0	■	46.3	16.1	15.6	9.9	13.5	11.3	16.4	14.6	6.0	7.8	29.0	9.2	10	NC_004306-Fibrovirus
11	487.0	524.0	325.0	386.0	1000.0	365.0	362.0	414.0	344.0	84.6	■	16.1	18.2	12.9	14.2	11.8	16.6	19.2	12.1	3.9	34.6	8.4	11	NC_005948-Villovirus
12	323.0	349.0	221.0	365.0	605.0	344.0	374.0	105.3	354.0	277.0	335.0	■	6.6	12.5	19.8	10.5	80.6	14.7	8.9	5.7	13.7	8.9	12	NC_006294-Capistrivirus
13	466.0	482.0	330.0	399.0	1000.0	360.0	374.0	374.0	456.0	273.0	260.0	418.0	■	10.3	15.0	13.4	10.0	33.7	11.1	6.6	20.8	12.7	13	NC_007189-Staminivirus
14	213.0	234.0	347.0	422.0	1000.0	403.0	354.0	344.0	386.0	471.0	374.0	292.0	434.0	■	11.8	23.6	10.6	11.3	34.9	10.6	14.5	20.8	14	NC_010429-Psecadovirus
15	487.0	517.0	168.5	426.0	1000.0	354.0	354.0	226.0	349.0	354.0	299.0	221.0	313.0	399.0	■	8.9	18.3	12.8	12.4	3.5	12.4	12.9	15	NC_015209-Affertcholeramvirus
16	144.7	163.8	342.0	426.0	719.0	447.0	456.0	365.0	451.0	451.0	493.0	323.0	511.0	209.0	383.0	■	8.4	10.5	26.5	8.9	13.2	26.4	16	NC_015297-Parhipatevirus
17	342.0	349.0	222.0	328.0	637.0	305.0	342.0	102.2	325.0	291.0	335.0	22.5	360.0	301.0	229.0	342.0	■	14.2	7.6	5.3	18.2	8.4	17	NC_016162-Vicialiavirus
18	524.0	649.0	313.0	451.0	1000.0	605.0	434.0	352.0	504.0	284.0	267.0	383.0	129.8	493.0	362.0	675.0	407.0	■	13.7	9.1	17.4	11.9	18	NC_021569-Subteminivirus
19	205.0	238.0	323.0	586.0	1000.0	498.0	577.0	476.0	487.0	447.0	399.0	430.0	418.0	115.5	396.0	155.8	399.0	482.0	■	7.6	11.6	25.1	19	NC_043028-Xylivirus
20	577.0	688.0	1000.0	1000.0	301.0	775.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	754.0	1000.0	1000.0	819.0	845.0	1000.0	1000.0	■	7.1	6.2	20	NC_055052-Habenivirus
21	393.0	442.0	309.0	344.0	524.0	371.0	371.0	335.0	374.0	148.2	126.5	319.0	244.0	383.0	301.0	386.0	319.0	252.0	368.0	1000.0	■	14.0	21	MZ474491 ●
22	149.5	198.0	383.0	430.0	874.0	414.0	498.0	360.0	430.0	352.0	399.0	357.0	447.0	209.0	418.0	158.4	342.0	553.0	179.7	736.0	399.0	■	22	MZ474490 ●
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		

Divergence (lower-left triangle)

b Percent Identity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1	■	20.5	11.4	25.0	20.5	9.1	13.6	6.8	18.2	18.2	27.3	18.2	15.9	20.5	13.6	1	NC_001418-Tertilicivirus
2	321.0	■	70.5	20.5	20.5	11.4	13.6	6.8	18.2	63.6	15.9	13.6	9.1	18.2	22.7	2	NC_004306-Fibrovirus
3	321.0	37.5	■	15.9	25.0	15.9	9.1	6.8	52.3	6.8	9.1	11.4	6.8	20.5	13.6	3	MZ474491 ●
4	335.0	775.0	775.0	■	45.2	17.7	14.5	6.5	45.2	22.6	6.5	25.8	12.9	45.2	9.7	4	MZ474488 ●
5	246.0	845.0	531.0	93.8	■	8.7	17.4	4.3	36.2	23.2	14.5	24.6	15.9	26.1	14.5	5	NC_015297-Parhipatevirus
6	661.0	234.0	264.0	206.0	305.0	■	57.5	8.2	23.3	26.0	9.6	17.8	27.4	16.4	9.6	6	NC_003287-Inovirus
7	357.0	234.0	234.0	129.8	277.0	61.8	■	6.8	16.2	8.1	13.5	17.6	37.8	8.1	14.9	7	NC_001954-Infulavirus
8	1000.0	1000.0	1000.0	447.0	626.0	1000.0	775.0	■	4.1	4.1	4.1	8.1	8.1	10.8	4.1	8	NC_015209-Affertcholeramvirus
9	410.0	399.0	1000.0	84.3	102.0	258.0	262.0	438.0	■	22.4	3.9	18.4	10.5	27.6	13.2	9	NC_043028-Xylivirus
10	430.0	49.4	74.0	301.0	305.0	305.0	309.0	1000.0	195.1	■	3.9	9.1	11.7	18.2	16.9	10	NC_005948-Villovirus
11	162.9	560.0	560.0	339.0	380.0	637.0	386.0	1000.0	504.0	403.0	■	8.9	8.9	6.3	5.1	11	NC_006294-Capistrivirus
12	586.0	328.0	399.0	135.4	160.9	299.0	205.0	357.0	165.4	386.0	719.0	■	23.2	18.3	4.9	12	NC_001331-Primolicivirus
13	303.0	271.0	271.0	177.9	257.0	123.6	87.4	1000.0	399.0	287.0	586.0	216.0	■	10.7	16.7	13	NC_001332-Lineavirus
14	252.0	386.0	321.0	88.9	160.9	254.0	191.1	517.0	123.0	208.0	434.0	213.0	279.0	■	13.5	14	NC_008575-Restivirus
15	430.0	215.0	271.0	217.0	261.0	273.0	257.0	362.0	204.0	253.0	422.0	233.0	231.0	347.0	■	15	NC_001396-Coriovirus
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		

Divergence (lower-left triangle)

Fig. 2 Sequence comparison of four novel inoviruses identified in this study. **a** Pairwise comparison of Zot amino acid sequences identified in this study with the representative strains of different genera of the family *Inoviridae*. **b** Pairwise comparison of CoaB amino acid sequences identified in this study with the representative strains of different genera of the family *Inoviridae*. Pairwise comparison was used Cluster W method which implemented into MegAlign program of DNAStar software. The "Percent Identity" and "Divergence" of sequences was calculated in default method and shown with picture
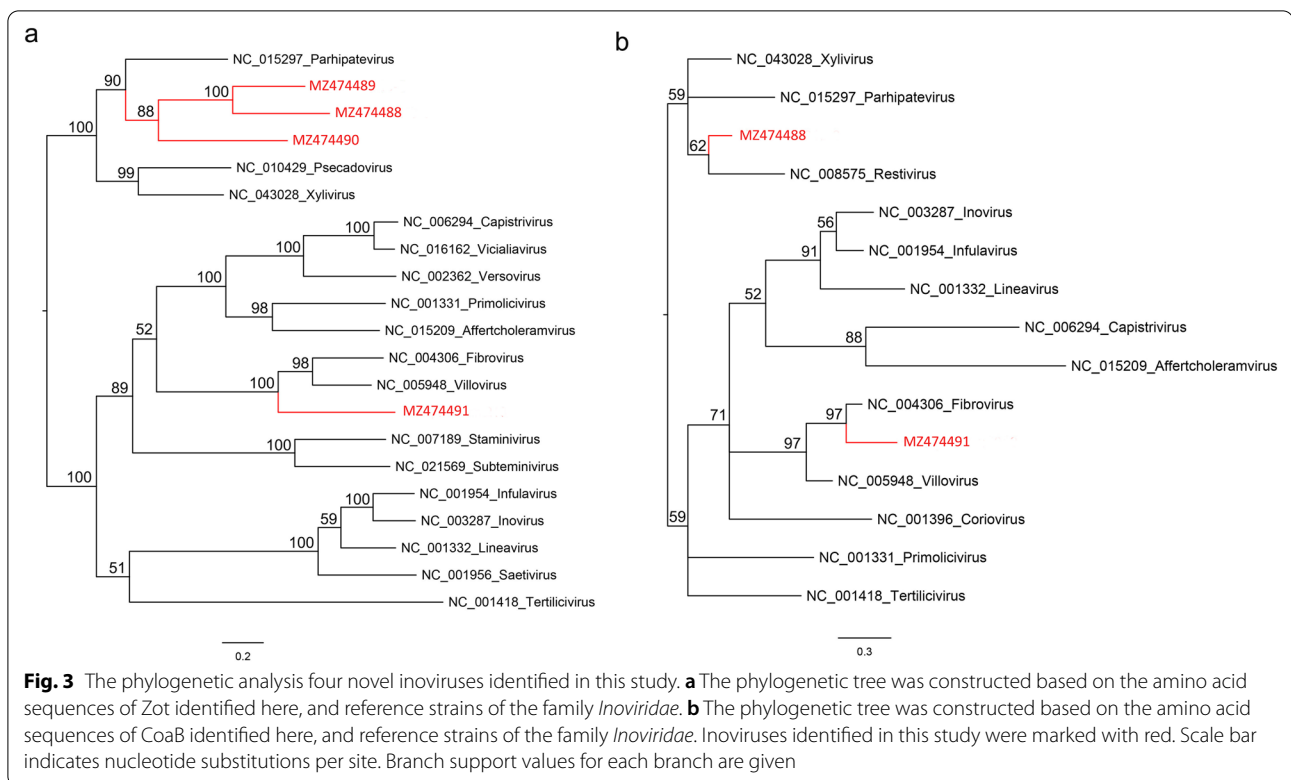
Peter Simmonds et al., proposal that genomics-based classification of metagenomically derived viruses should be incorporated into the ICTV taxonomy in the future [20, 21].

Two phylogenetic trees were constructed based on Zot and CoaB proteins including the reference strains of different genera of the family *Inoviridae* (Fig. 3). The result showed that all four inoviruses were forming an independent clade in the Zot tree separately (Fig. 3a). The strains MZ474489, MZ474488, and MZ474490 had a closer genetic distance with the strain NC_015297 of genus *Parhipatevirus*, while the strain MZ474491 had a closer genetic distance with the strain NC_005948 of genus *Villovirus*. Because the CoaB proteins of MZ474489 and MZ474490 failed to identify in this study, only the MZ474488 and MZ474491 showed in the CoaB tree. The strain MZ474488 had a closer genetic distance with the strain NC_008575 of genus *Restivirus*, but the strain MZ474491 had a closer genetic distance with the strain NC_004306 of genus *Fibrovirus* (Fig. 3b).

Conclusions

In summary, we detected four novel filamentous viruses in fecal samples of wild and breeding birds and characterized their complete genome. All four inoviruses encoded the Zot proteins which were conserve in the members of the family *Inoviridae* and functioned as a key role in increasing the pathogenicity of lysogenic

Zeng *et al. Virology Journal*   (2021) 18:240

Page 5 of 6

Fig. 3 The phylogenetic analysis four novel inoviruses identified in this study. **a** The phylogenetic tree was constructed based on the amino acid sequences of Zot identified here, and reference strains of the family *Inoviridae*. **b** The phylogenetic tree was constructed based on the amino acid sequences of CoaB identified here, and reference strains of the family *Inoviridae*. Inoviruses identified in this study were marked with red. Scale bar indicates nucleotide substitutions per site. Branch support values for each branch are given

bacteria. Homology comparison based on the Zot proteins indicated that all those four inoviruses shared the highest amino acid sequence identity (< 37.5%) to other reference strains belonging to the family *Inoviridae* and among themselves. Phylogenetic analysis showed that all four inoviruses in this study were far away from other reference strains of different genera of the family *Inoviridae*. According to the genetic distance-based criteria, all four inoviruses belonged to four novel genera of the family *Inoviridae*. This study proved that viral metagenomics approach was suitable for the exploration and identification of filamentous phages.

Abbreviations
Zot: Zonula occludens toxin; CoaB: Major coat protein; Rep: Replication protein; DPBS: Dulbecco's phosphate buffered saline; ORF: Open reading frame; BLAST: Basic local alignment search tool.

Authors' contributions
WZ conceived the study. JZ and YW performed all the experiments. JZ, YW, SY, and JZ analyzed the data. WZ wrote the paper. All authors read and approved the final manuscript.

Declarations

Ethics approval and consent to participate
Ethical Approvals were given by the Ethics Committee of Chinese Academy of Agricultural Sciences with the reference number of SVRI2017091.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Faruque SM, Bin Naser I, Fujihara K, Diraphat P, Chowdhury N, Kamruzzaman M, et al. Genomic sequence and receptor for the Vibrio cholerae phage KSF-1phi: evolutionary divergence among filamentous vibriophages mediating lateral gene transfer. J Bacteriol. 2005;187:4095–103.
2. Kuo TT, Tan MS, Su MT, Yang MK. Complete nucleotide sequence of filamentous phage Cf1c from *Xanthomonas campestris* pv. citri. Nucleic Acids Res. 1991;19:2498.

Zeng *et al. Virology Journal*   (2021) 18:240

Page 6 of 6

3. Ehara M, Shimodori S, Kojima F, Ichinose Y, Hirayama T, Albert MJ, et al. Characterization of filamentous phages of Vibrio cholerae O139 and O1. FEMS Microbiol Lett. 1997;154:293–301.
4. Murugaiyan S, Bae JY, Wu J, Lee SD, Um HY, Choi HK, et al. Characterization of filamentous bacteriophage PE226 infecting Ralstonia solanacearum strains. J Appl Microbiol. 2011;110:296–303.
5. Zinder ND, Valentine RC, Roger M, Stoeckenius W. F1, A ROD-shaped male-specific bacteriophage that contains DNA. Virology. 1963;20:638–40.
6. Messing J. M13 cloning vehicles. Their contribution to DNA sequencing. Methods Mol Biol. 1993;23:9–22.
7. Nissim A, Hoogenboom HR, Tomlinson IM, Flynn G, Midgley C, Lane D, et al. Antibody fragments a "single pot" phage display library as immunochemical reagents. EMBO J. 1994;13:692–8.
8. Knezevic P, Voet M, Lavigne R. Prevalence of Pf1-like (pro)phage genetic elements among aeruginosa isolates. Virology. 2015;483:64–71.
9. Secor PR, Sweere JM, Michaels LA, Malkovskiy AV, Lazzareschi D, Katznelson E, et al. Filamentous bacteriophage promote biofilm assembly and function. Cell Host Microbe. 2015;18:549–59.
10. Burgener EB, Sweere JM, Bach MS, Secor PR, Haddock N, Jennings LK, et al. Filamentous bacteriophages are associated with chronic Pseudomonas lung infections and antibiotic resistance in cystic fibrosis. Sci Transl Med. 2019. https://doi.org/10.1126/scitranslmed.aau9748.
11. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019;4:1895–906.
12. Tisza MJ, Buck CB. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. Proc Natl Acad Sci U S A. 2021;118:e2023202118.
13. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1:18.
14. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
15. Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 2013;41:W22–8.
16. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
17. Prangishvili D, Bamford DH, Forterre P, Iranzo J, Koonin EV, Krupovic M. The enigmatic archaeal virosphere. Nat Rev Microbiol. 2017;15:724–39.
18. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV. Viruses of archaea: structural, functional, environmental and evolutionary genomics. Virus Res. 2018;244:181–93.
19. Yutin N, Bäckström D, Ettema TJG, Krupovic M, Koonin EV. Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. Virol J. 2018;15:67.
20. Iranzo J, Koonin EV, Prangishvili D, Krupovic M. Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. J Virol. 2016;90:11043–55.
21. Aiewsakun P, Adriaenssens EM, Lavigne R, Kropinski AM, Simmonds P. Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. J Gen Virol. 2018;99:1331–43.

Publisher's Note