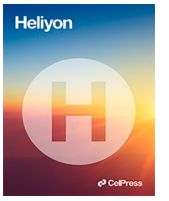
Heliyon 10 (2024) e38338

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Identification and characterization of multiple novel viruses in fecal samples of ruddy shelducks using viral metagenomics methods

Li Ji [a,b,1], Yan Wang [a,1], Yijie Sun [a,1], Likai Ji [a], Xiaochun Wang [a], Yuwei Liu [a], Quan Shen [a], Shixing Yang [a,*], Wen Zhang [a,**]

[a] Department of Microbiology, School of Medicine, Jiangsu University, Zhenjiang, Jiangsu, 212013, China
[b] Zhenjiang Mental Health Center, Zhenjiang, Jiangsu, 212005, China

ARTICLE INFO

Keywords:
Viral metagenomics
Ruddy shelduck
Genomic structure
Phylogenetic analysis

ABSTRACT

The viral metagenomics approach is an effective technique for investigating and analysing both existing and emerging viruses in humans and diverse animal samples. The ruddy shelduck, a nationally protected secondary key species of wild animals, has become the predominant species among overwintering waterbirds in Qinghai Lake. Viruses carried by ruddy shelducks can potentially infect humans or other animals; however, limited research on the faecal virome of ruddy shelducks is currently available. In the present study, faecal samples of ruddy shelducks collected from Saga County, Shigatse City, Tibet, China, were subjected to viral metagenomic analysis. The predominant viral families identified in ruddy shelduck samples were *Picornaviridae*, *Parvoviridae*, *Microviridae*, *Vilyaviridae, Astroviridae*, and *Caliciviridae*. Among these, two picornavirus genomes have been identified as new strains of the genus *Megrivirus* in the family *Parvoviridae*. In addition, viruses that infect parasites and bacteria have been identified and characterised. The present study enhances our comprehension of the composition of the viral community in ruddy shelducks faeces and highlights the dynamic nature of viral evolution and the significance of continuous monitoring to assess potential risks to wildlife and public health.

1. Introduction

The ruddy shelduck (*Tadorna ferruginea*), classified as a nationally protected secondary key wild animal in China, has historically exhibited high abundance within its range during winter and migration periods across northern China. In addition, prevalent within the lower Yangtze River region and highly regarded within China's rare bird breeding industry, this species has been recognized for its popularity among breeders [1]. The Qinghai-Tibet Plateau is a primary wintering ground because of its freshwater habitats, that attract substantial numbers of migratory individuals originating from their respective breeding grounds elsewhere. It is an omnivorous avian species with dietary preferences encompassing fish, shrimp, crustaceans such as crabs, and aquatic plants, including crop remnants. Currently, the number of ruddy shelducks is declining because of excessive hunting practices, coupled with extensive habitat

* Corresponding author.
** Corresponding author.
E-mail addresses: 1000004113@ujs.edu.cn (S. Yang), zhangwen@ujs.edu.cn (W. Zhang).
[1] These authors have equally contributed to this work.

https://doi.org/10.1016/j.heliyon.2024.e38338
Received 15 April 2024; Received in revised form 21 September 2024; Accepted 23 September 2024
Available online 26 September 2024

degradation [2]. The global population is estimated to be approximately 30,000, with the vast majority living in Asia, including China, India, Nepal, Pakistan, and Myanmar. In addition, viruses that infect birds and other ducks can endanger their lives. Recently, ducks have shown susceptibility to Tambusu virus (TMUV), West Nile virus (WNV), duck hepatitis virus (DHV), highly pathogenic avian influenza virus (HPAIV), and other emerging pathogens that pose a substantial risk to public health [3–7]. Moreover, there is a possibility of interspecies transmission of these viruses with zoonotic implications. However, from a logical perspective, ruddy shelducks remain understudied.

The viral metagenomics approach serves as a potent technique for exploring both new and recognized viruses, and it is widely used to analyze the viral composition across a variety of samples [8–11]. Nevertheless, there is currently limited research on the viral components present in ruddy shelducks faecal samples. In this study, we employed a viral metagenomics approach to investigate the viral composition of ruddy shelducks fecal samples collected from Saga County, Shigatse City, Tibet Autonomous Region, China. The findings of this study contribute to our understanding of the virome of ruddy shelducks and offer valuable information into viral genomes, which are essential for investigating the viromes of other avian species within the same region.

2. Materials and methods

2.1. Sample collection and preparation

A total of 10 faecal samples from 10 healthy ruddy shelducks were collected in December 2020 from Saga County, Shigatse City, Tibet Autonomous Region, China. Disinfect cotton swabs were used to obtain faecal samples, which were stored in sterile tubes and cold chain transportation to the laboratory. Each fecal sample was mixed with 1.5 ml of phosphate-buffered saline (PBS), vigorously vortexed for 10 min, and then centrifuged at 12,000 rpm for 10 min. The supernatant from each sample was then transferred to a new 1.5-ml centrifuge tube and stored at −80 °C for the following study.

2.2. Viral nucleic acid extraction

A combined supernatant sample pool containing ten faecal supernatants was generated. A total of 500 μl faecal suspension (50 μl faecal suspension per faecal sample) was mixed and filtered through a 0.45-μm filter (Merck Millipore, Billerica, MA, USA) centrifuging at 8000 rpm for 5 min to eliminate bacteria and particles of eukaryotic cell size. Subsequently, the virus-rich filtrates were treated at 37 °C for 90 min with a mixed nuclease cocktail (Turbo DNase from Thermo Fisher Scientific, USA, BaselineZero DNase from Epicentre, USA, Benzoase nuclease from Novagen Corporation, USA, and RNase A from Thermo Fisher Scientific) to degrade unprotected nucleic acid [12–14]. According to the manufacturer's protocol, the QIAamp MinElute Virus Spin Kit (Qiagen, Hilden, Germany) was employed to extract viral RNA and DNA. The concentrations of DNA and RNA were determined using the Qubit™4 Nucleic Acid Concentration Sequencer (Invitrogen, Carlsbad, CA, USA).

2.3. Library construction and bioinformatics analysis

The cDNA of viruses were synthesized by reverse transcription procedure using random hexamer primers. Complementary chains of cDNA were synthesized using Klenow fragment DNA polymerase (New England Biolabs, Ipswich, MA, USA). Subsequently, a library was constructed using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA), followed by sequencing on the NovaSeq Illumina platform with paired-end reads of 250 bp and dual barcoding for each pool.

For bioinformatics analysis, 250-bp paired-end reads generated by NovaSeq were debarcoded using vendor software, which was processed using an in-house analysis pipeline running on a 32-node Linux cluster. Low-quality sequenced tails were trimmed using a Phred quality score of 10 as the threshold. Adapter sequences were removed utilising the default VecScreen settings. Bowtie2 v2.2.4 was employed to align bacterial reads against the bacterial nucleotide sequences in the Basic Local Alignment Search Tool (BLAST) NT database, effectively subtracting bacterial contamination from the dataset. *De novo* reassemble the clean reads using the default setting (Kmer size 63) of SOAPdenovo2 version r240 [15]. The assembled contigs and singlets were compared against an in-house viral proteome database using BLASTx (v.2.2.7), with an E-value cutoff $<10^{-5}$. To obtain longer contigs or the complete genome, each viral contig was employed as a reference, and the original data were mapped to it using the Low Sensitivity/Fastest setting in Geneious 11.1.2.

2.4. Phylogenetic analysis

Phylogenetic analysis was performed based on the amino acid sequences of viruses identified in this study, their closest BLASTx matches in GenBank, and the representative members of relevant viral species or genera. Sequence alignment was performed using the MUSCLE algorithm with default settings in MEGA-X [16], and ambiguously aligned regions and gaps were manually stripped using Geneious 11.1.2. Phylogenetic trees were constructed using MrBayes version 3.2.7. The parameter "prset aamodelpr = mixed" was used to allow the program to utilize the ten built-in amino acid substitution models. The analysis was set to run for a maximum of ten million generations, with sampling every 50 generations. The first 25 % of the Markov Chain Monte Carlo (MCMC) samples were discarded as part of the burn-in process. Convergence was confirmed when the standard deviation of split frequencies dropped below 0.01. Bootstrap values were assigned to each node in the tree. The National Center for Biotechnology Information (NCBI) Open Reading Frame (ORF) Finder and Geneious 11.1.2. were used to predict putative open reading frames (ORFs).

2.5. Nucleotide sequence accession number

The original sequence data analyzed in our study have been deposited in the Genome Sequence Archive (Genomics, Proteomics, and Bioinformatics 2021) at the National Genomics Data Centre (Nucleic Acids Res 2021), China National Centre for Bioinformation, Beijing Institute of Genomics, and Chinese Academy of Sciences (GSA: CRA014768). It is also publicly available at https://ngdc.cncb.ac.cn/gsa. gsa.doc.faq.h4CB1. The genome reported in the current study has been stored in GenBank at the National Genomics Data Centre, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Centre for Bioinformation under accession numbers C_AA062773.1–C_AA062778.1, which are accessible to the public: https://ngdc.cncb.ac.cn/genbase.

3. Results

3.1. Overview of viral metagenome

The library was sequenced on an Illumina NovaSeq platform to generate 437,794 raw sequence reads with an average percentage of GC content (GC %) of 54.4 %. Among them, 13,211 sequence reads had the best match with viral proteins, accounting for 3.0 % of the raw data reads. Subsequent analyses identified the distribution of viral sequence reads across the different viral families. The majority of sequence reads among the families were assigned to the *Picornaviridae* family (83.6 %), followed by those *Parvoviridae* (12.4 %) and *Microviridae* (2.5 %). The remaining viral families included *Vilyaviridae* (0.7 %), *Astroviridae* (0.5 %), and *Caliciviridae* (0.4 %), each representing a smaller proportion (Fig. 1). Viruses from families such as *Parvoviridae*, *Astroviridae*, and *Caliciviridae* were omitted because their contigs were too short to provide useful information. Virus from the families *Picornaviridae, Microviridae*, and *Vilyaviridae* were further detailed analyzed.

3.2. Two picornaviruses belonging to the Megrivirus genus of the Picornaviridae family

A total of 7771 raw sequence reads belonging to the *Picornaviridae* family were identified in this library, of which 6378 were successfully mapped to the same picornavirus. Two almost complete genomes of picornaviruses were obtained using the "Map to Reference" program in Geneious 11.1.2. BLASTn analyses indicated that these two picornaviruses were closely associated with members of the *Megrivirus* genus, such as megriviruses found in swans in the UK (MW588051 and MW588050) and were named XZDM1 (mean coverage: 289.6) and XZDM2 (mean coverage: 301.2), respectively. The genomes of XZDM1 and XZDM2 were 9356 and 9694 nucleotides (nt) in length, with an percentage of GC content of 45.5 % and 45.3 %, respectively. Both XZDM1 and XZDM2 possessed a single ORF that encoded polyproteins comprising 2843 and 2848 amino acids (aa), respectively, with untranslated regions (UTR) of different lengths at both ends. Similar to other picornaviruses, the polyproteins of XZDM1 and XZDM2 can undergo cleavage resulting in the generation of VP0, VP3, VP1, hypothetical protein, 2A-2C, and 3A-3D by comparison with the polyproteins of the chicken picornavirus 4 strain found in chicken in China (NC_024768) and the megrivirus C2 strain found in chicken in China (NC_024769) (Fig. 2a). The P1 polypeptides of XZDM1 and XZDM2 consist of 1177 aa and 1095 aa and were cleaved at the VP0/VP3 (E^{421}/G^{422} and T^{427}/G^{428}), VP3/VP1 (Q^{587}/G^{588} and Q^{592}/S^{593}), and VP1/hypothetical protein (Q^{847}/A^{848} and Q^{852}/A^{853}). They had

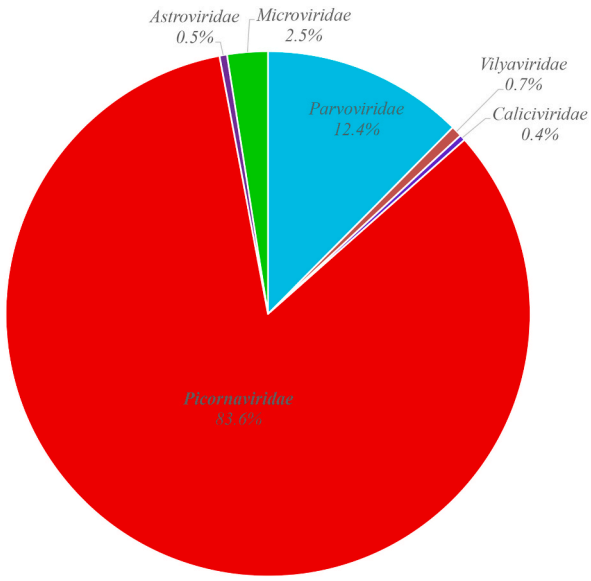


Fig. 1. The composition of fecal virome detected in ruddy shelduck. The percentage of virus sequences in different virus family were shown in pie chart.

the highest aa identities of 94.66 % and 96.53 % with the *Riboviria* sp. Isolate found in birds in China (OQ425090) and Megrivirus A strain found in swans in the UK (MW588050), respectively. The P2 polypeptides of XZDM1 and XZDM2, comprising 868 and 869 aa, respectively, contain three nonstructural proteins: 2A (cleavage sites: Q^{1446}/S^{1447} and Q^{1365}/S^{1366}), 2B (cleavage sites: Q^{1661}/S^{1662} and Q^{1580}/S^{1581}), and 2C. The BLASTp results revealed that the P2 regions of XZDM1 and XZDM2 had the highest aa identities of 97.12 % and 97.81 %, respectively, with that of the *Picornaviridae* sp. isolate found in Anser albifrons in China (MT138342). A conserved Walker A motif exists in the 2C protein region (Fig. 2a). The P3 polypeptides of XZDM1 and XZDM2 are 798 and 884 amino acids, respectively, and are cleaved into four non-structural proteins, 3A, 3 B, $3C^{pro}$ (protease), and 3D, at sites 3A/3B (E^{2224}/A^{2225} and E^{2143}/A^{2144}), 3B/3C (E^{2253}/G^{2254} and E^{2172}/G^{2173}), and 3C/3D (E^{2455}/A^{2456} and E^{2374}/A^{2375}). The P3 regions of XZDM1 and XZDM2 shared the highest aa identities (98.60 % and 98.8 7 %, respectively) with the megrivirus A2 strain found in ducks in China (NC_024120). The presence of conserved polymerase and proteinase motifs, such as YGDD, KDE, and GXCGX10-15GXH, was observed in the 3C and 3D proteins of XZDM1 and XZDM2, whereas the conserved FLKR polymerase motif was only found in the 3D protein of XZDM2 and not in XZDM1 (Fig. 2a).

The P1 and 3CD regions of the representative strains belonging to the *Megrivirus* genus of the *Picornaviridae* family, as well as the XZDM1 and XZDM2, were subjected to phylogenetic analysis (Fig. 2b and c). In the P1 phylogenetic tree, XZDM1 clustered with the Megrivirus A strain MK204391 identified from faecal samples of Pacific black ducks in Australia in 2012, forming a separate branch, whereas XZDM2 clustered with other Megrivirus A strains (KC663628, MK204417, KY369299, MW588054, and MW588050) isolated from faecal samples of ducks, pink-eared ducks, geese, and Cygnus oleracea, forming another branch. The results of the 3CD phylogenetic tree showed that XZDM1 and XZDM2 clustered with other Megrivirus A strains, forming a branch. Based on these results, the

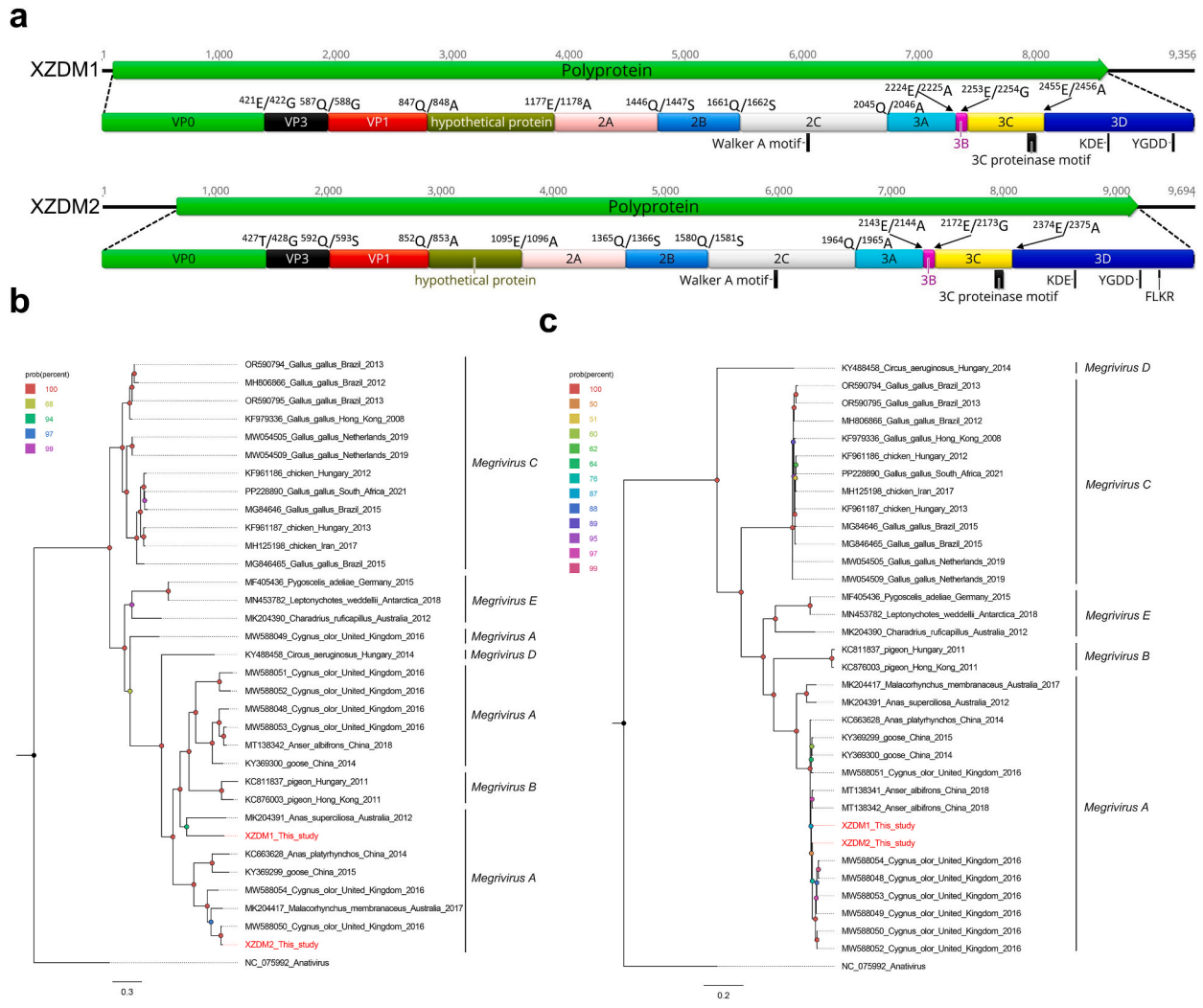


Fig. 2. The genomic organization, conserved motifs, and phylogenetic analysis of these megriviruses identified in ruddy shelduck. (a) The genomic organization of two megrivirus strains. The ORFs and viral encoding proteins of megriviruses were marked with different colors. The conserved motifs were also shown. (b), (c) The phylogenetic analysis based on the P1 region and 3CD of megriviruses, which identified in this study, and other reference strains belonging to the *Megrivirus* genus of the *Picornaviridae* family. XZDM1 and XZDM2 identified in this study were marked with red.

3CD region of megriviruses is considered to be more suitable for studying viral evolution. According to the International Committee on Taxonomy of Viruses (ICTV) classification for picornaviruses (https://ictv.global/report/chapter/picornaviridae/megrivirus) (accessed on January 28th, 2024), members of a picornavirus species should exhibit significant aa identity in their P1, 2C, 3C, and 3D proteins. Specifically, for members of the genus *Megrivirus*, there should be <55 % difference in P1 amino acid identity and <40 % difference in the combined amino acid identity between regions 2C+3CD. Based on this classification standard, XZDM1 and XZDM2 were classified as *Megrivirus A* species.

3.3. A single-stranded DNA virus belonging to the Herugrimvirus genus of the Vilyaviridae family

A total of 63 raw sequence reads belonging to the *Vilyaviridae* family were identified in this library. One complete genome of Vilyavirus was obtained using the "Map to Reference" program in Geneous 11.1.2 and designated as XZCO1 (mean coverage: 55.3). The genome of XZCO1 was 2092 nt long and contained two bidirectional ORFs encoding a replicase-associated protein (Rep) and a capsid protein (CP) (Fig. 3a). The CP protein comprises 259 amino acids and is the primary structural protein of this virus. The Rep protein is 335 aa in length and plays a crucial role in viral replication. The Rep protein of XZCO1 lacked conserved rolling-circle replication (RCR) motifs I, II, and III, whereas conserved superfamily 3 (SF3) helicase motifs (Walker A: GPSGTGKS, Walker B: VLDDF, and Walker C: LTSN) were present.

Phylogenetic analysis were performed based on the Rep proteins showed that XZCO1 clustered with other members of the *Herugrimvirus* genus, forming an independent branch (Fig. 3b). The BLASTn results showed that XZCO1 shared nucleotide sequence identities of 97.64 %, 97.64 %, and 97.90 % with strains MT293421, MT293422, and MT293423 found in giardia in Netherlands, respectively, and only 84.57 % nucleotide sequence identity with another strain (MT293424) of *Herugrimvirus gladden*. According to the classification standard of the *Vilyaviridae* family proposed by Krupovic, The new virus must be classified within a species if it shares more than 78 % nucleotide sequence identity with any variant already classified as belonging to that species [17]. Therefore, XZCO1 should be classified as a *Herugrimvirus gladden* species.

3.4. Three novel microviruses belonging to the Microviridae family

A total of 233 raw sequence reads belonging to the *Microviridae* family were identified in this library, and all sequence reads were matched to microviruses. Three complete genomes of microviruses were successfully obtained using the "Map to Reference" program in Geneious 11.1.2. one microvirus was related (38.2 %) to other members of the *Chlamydiamicrovirus* genus, whereas the other two microviruses had 66.6 % and 75.2 % nucleotide sequence identities with two unclassified microviruses [18]. Consequently, they were designated XZMCV1 (mean coverage: 13.4), XZMCV2 (mean coverage: 10.0), and XZMCV3 (mean coverage: 35.6). The genomes of XZMCV1, XZMCV2, and XZMCV3 were 4447 nt, 5197 nt, and 5571 nt, respectively, and contained five ORFs encoding different proteins. Specifically, both XZMCV3 and XZMCV2 encode major capsids, minor capsids, replication initiators, and scaffold proteins within their genomes. Furthermore, XZMCV3 contained one ORF encoding a peptidase, whereas XZMCV2 harboured one ORF responsible for DNA-binding protein synthesis. Among the five proteins encoded by XZMCV1, there are two proteins with unknown

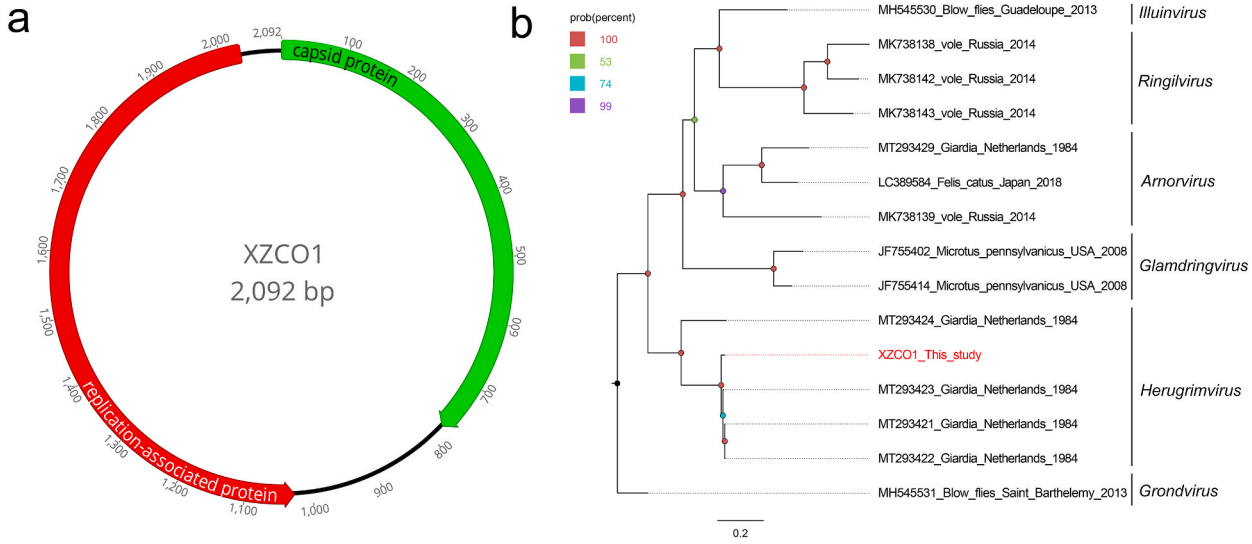


Fig. 3. The genomic organization and phylogenetic analysis of herugrimvirus identified in ruddy shelduck. (a) The genomic organization of one herugrimvirus identified in ruddy shelduck. Viral encoding proteins of XZCO1 were marked with different colors. (b) The phylogenetic analysis based on the Rep protein of XZCO1 which identified in this study, and reference strains of other members in the *Vilyaviridae* family. The XZCO1 identified in this study was marked with red.

functions named hypothetical proteins; the remaining three proteins are a major capsid protein, a DNA pilot protein and a replication initiator protein (Fig. 4a).

Phylogenetic tree was constructed based on the major capsid proteins of XZMCV1, XZMCV2, XZMCV3, and other representative strains belonging to various genera of the *Gokushovirinae* subfamily. These results revealed that the three microviruses were positioned in distinct branches. Among them, XZMCV3 clustered with two unidentified viruses identified from the human metagenome datasets in the USA (BK028880 and BK024429) [19]. XZMCV2 was grouped with two unidentified viruses isolated from tracheal swabs of *Gallus gallus* in USA (MN379630 and MN379640). XZMCV1 formed an independent branch cluster with other strains of the *Chlamydiamicrovirus* genus identified in different environments and biological samples, such as lake water, robins, Apis mellifera ligustica, and Tramea lacerata (MW697647, MZ364279, MH992208, and JX185431) [20] (Fig. 4b).

4. Discussion

Ducks, often referred to as "Trojan horses," can shed and transmit disease-causing viruses without exhibiting noticeable signs of illness, thereby facilitating cross-species transmission. Among species susceptible to viral infections, ruddy shelducks have been identified as particularly vulnerable. Although previous studies have documented various viruses capable of infecting ruddy shelducks and causing diseases [21–23], little is currently known regarding the faecal virome of ruddy shelducks. This study employed high-througtput sequencing technology to investigate the viral composition in ruddy shelducks faeces. Our findings represent the first identification of various viruses in ruddy shelduck faeces.

The faecal virome composition of ruddy shelducks in this study primarily included the families *Picornaviridae*, *Parvoviridae*, *Microviridae*, *Vilyaviridae*, *Astroviridae*, and *Caliciviridae*, similar to the virome composition in the Wille et al. study [11]. This is significantly different from the result of the previous Raimirez-Martinez et al. study. The fecal virome composition of the migratory wild duck species in their study consisted of viruses from the families *Myoviridae*, *Retroviridae*, *Adenoviridae*, *Alloherpesviridae*, and *Herpesviridae* [24]. The differences in virome composition may be due to differences in the host species, factors in different living environments, or methods for library construction.

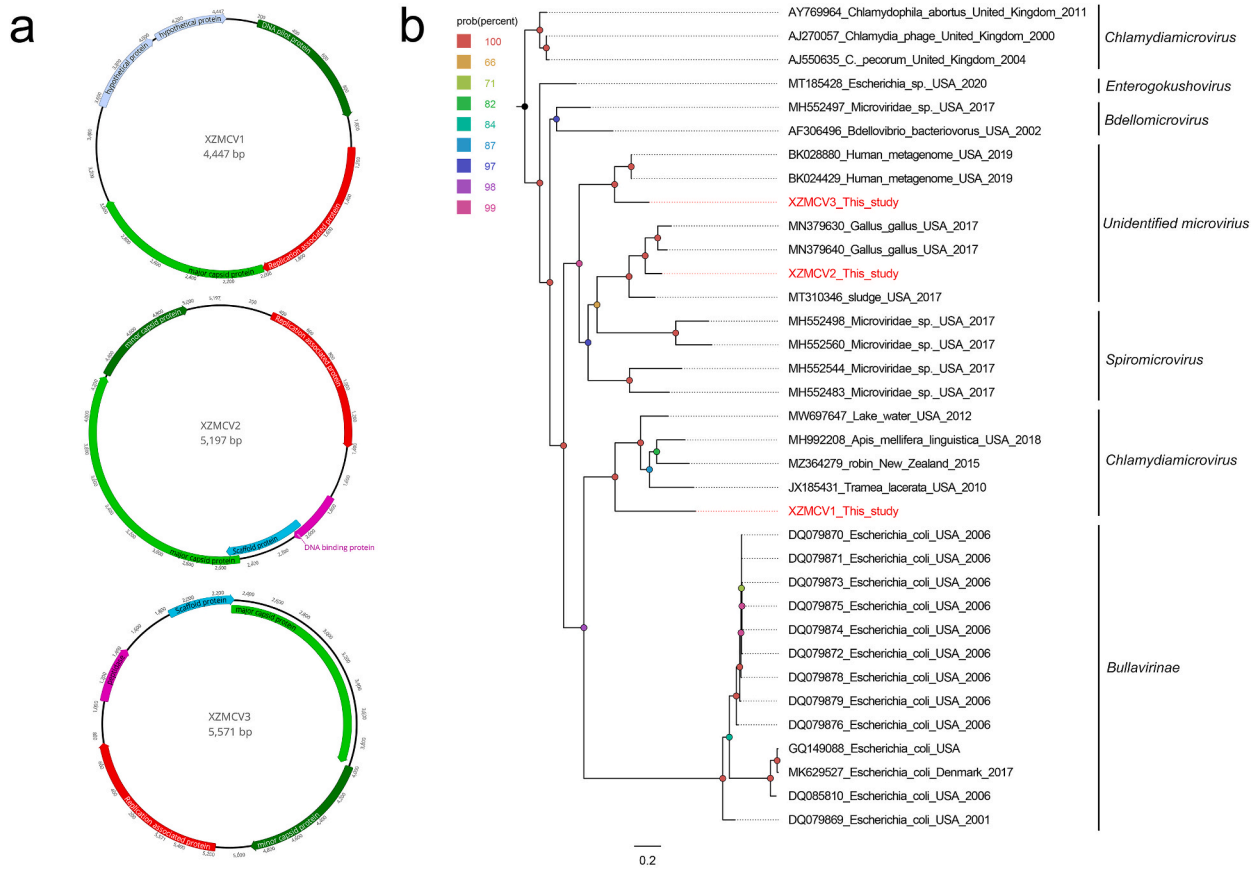


Fig. 4. The genomic organization and phylogenetic analysis of microviruses identified in ruddy shelduck. (a) The genomic organization of three microviruses identified in ruddy shelduck. Viral encoding proteins of three microviruses were marked with different colors. (b) The phylogenetic analysis based on the major capsid proteins of three microviruses which identified in this study, and reference strains of other members in the *Gokushovirinae* subfamily. XZMCV1, XZMCV2, and XZCV3 identified in this study were marked with red.

Members of the *Picornaviridae* family are small single-stranded RNA viruses with genome lengths ranging from approximately 7.2 to 9.4 kb. This family encompasses 158 species distributed across 68 genera. Many genera within the *Picornaviridae* family possess zoonotic potential, infecting both animals and humans and causing a diverse range of diseases [25–29]. Recently, picornavirus has gained significant attention within the public health community because of its seasonal variation in abundance and quantity, co-infection with other viruses or multiple picornavirus strains, and the potential emergence of new viral strains through recombination. Despite *Megrivirus* not being a new genus within the *Picornaviridae* family, our study identified two novel strains of Megrivirus A in faecal samples obtained from healthy ruddy shelducks. Through aa sequence comparison, we elucidated that the P1 of XZDM1 had an aa sequence identity of 52.72 % with that of the Megrivirus A strain found in swans in the UK (MW588050), whereas the P1 of XZDM2 had the highest aa sequence identity (96.53 %). This variation may suggest different selective pressures acting on structural proteins, which could be related to host adaptation or virus-host interactions. The consistent stability in the nonstructural region suggests that the 3CD protein may serve as a more reliable basis for classifying and studying the evolutionary relationships among megriviruses.

Vilyaviruses have single-stranded DNA genomes that are approximately 1.9–2.2 kb in length. Recently, Kinsella et al. identified them as being associated with protozoan parasites of the genus *Giardia* using viral megagenomics combined with statistics and experiments [30]. The XZCO1 strain identified here shared nucleotide sequence identities of 97.64 %, 97.64 %, and 97.90 %, respectively, with strains MT293421, MT293422, and MT293423, which were isolated from human stool and used *Giardia duodenalis* as a host; therefore, we believe that XZCO1 originated from parasites inside ruddy shelducks. To verify our hypothesis, we aligned the genome sequence of XZCO1 to whole-genome shotgun contigs of *Giardia* using BLASTn at NCBI. The results showed that the viral query aligned with high identity (95.31 %) and coverage (68 %) of nucleotide from *Giardia* genomes, suggesting *Gifardia* is the host of strain XZCO1. Further parasite isolation and microscopic morphological observations will be beneficial for identifying the true host. Surprisingly, unlike members of the *Circoviridae* family, the conserved RCR motifs I, II, and III were lacking in the replicase-associated protein of XZCO1 and other herugrimviruses, suggesting that an alternative replication mechanism was employed. However, further functional experiments are required to validate this hypothesis.

Gokushoviruses are circular, single-stranded DNA bacteriophages with a genome length ranging from 4.0 to 6.0 kb, composed of three to 11 genes [31]. Gokushoviruses have been primarily detected in diverse environments and biological samples, including marine environments, peatlands, methane seep sediments, dragonflies, humans, turkeys, and bats [32]. Although gokushoviruses are widespread in the environment, the only ones that have been isolated are lytic parasites obtained from host-restricted intracellular bacteria such as *Spiroplasma*, *Chlamydia*, and *Bdellovibrio* [33]. However, the hosts of most gokushoviruses are unknown. In this study, three novel gokushoviruses were identified in the faecal samples of ruddy shelducks. These viruses have low nucleotide sequence identity with other members of the *Chlamydiamicrovirus* genus in the *Gokushoviridae* family and unclassified microviruses. The characterisation of the three novel microviruses, each with a unique genomic architecture and protein composition, expands our understanding of the *Microviridae* family. Because gokushoviruses are widely dispersed in the natural environment, we could not determine whether the three gokushoviruses in this study came from ruddy shelducks or from external contamination. Environmental sampling for virus detection will help us further determine their sources.

Despite the significant insights provided by our study regarding the viral landscape of ruddy shelducks, several limitations must be acknowledged. First, the sample size was relatively small, potentially limiting the representativeness of our findings. A larger sample size would enhance the statistical power of our results and improve our understanding of the prevalence and diversity of viruses in this bird species. Second, although high-throughput sequencing technologies have allowed the identification of a wide array of viruses, their functional relevance remains uncertain. It is unclear whether all detected viruses actively replicate and contribute to disease processes, or if some are transiently present due to environmental exposure [11]. Further studies are required to characterise the pathogenic potential and replication status of these viruses. Furthermore, the origins of some viruses, particularly bacteriophages, such as gokushoviruses, could not be conclusively determined. Without detailed environmental sampling and control groups, it is challenging to discern whether these viruses originated from the birds themselves or from external sources. Finally, the complex nature of viral-host interactions and potential coinfections add another layer of complexity to interpreting virome data. The present study did not fully explore the potential synergistic effects of co-infections or the impact of host ecology on the structure of the virome [34]. Future studies should consider these factors to better understand how viruses interact within and between host species.

5. Conclusions

In conclusion, the findings of this study contribute to a better understanding of the virome in ruddy shelducks and highlight the need for ongoing surveillance and research on the role of these birds in the ecology of emerging viruses. The discovery of novel viruses, particularly within the *Picornaviridae*, *Vilyaviridae*, and *Microviridae* families, emphasises the dynamic nature of viral evolution and the importance of continuous monitoring to assess the potential risks to wildlife and public health.

Funding statement

This research was supported by the National Key Research and Development Programs of China (2023YFD1801300) and the National Natural Science Foundation of China (82341106).

Ethics statements

This study was reviewed and approved by Jiangsu University Ethics Committee with the approval number: 2018ujs18023 on the use of animals and complied with Chinese ethics laws and regulations.

Data availability statement

The original sequence data analyzed in our study have been deposited in the Genome Sequence Archive (Genomics, Proteomics, and Bioinformatics 2021) at the National Genomics Data Centre (Nucleic Acids Res 2021), China National Centre for Bioinformation, Beijing Institute of Genomics, and Chinese Academy of Sciences (GSA: CRA014768). It is also publicly available at https://ngdc.cncb.ac.cn/gsa. gsa.doc.faq.h4CB1. The genome reported in the current study has been stored in GenBank at the National Genomics Data Centre, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Centre for Bioinformation under accession numbers C_AA062773.1–C_AA062778.1, which are accessible to the public: https://ngdc.cncb.ac.cn/genbase.

CRediT authorship contribution statement

Li Ji: Writing – original draft, Formal analysis, Data curation. **Yan Wang:** Formal analysis, Data curation. **Yijie Sun:** Formal analysis, Data curation. **Likai Ji:** Formal analysis, Data curation. **Xiaochun Wang:** Formal analysis, Data curation. **Yuwei Liu:** Formal analysis, Data curation. **Quan Shen:** Formal analysis, Data curation. **Shixing Yang:** Conceptualization. **Wen Zhang:** Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Wang, S. Zheng, L. Li, Y. Yang, Y. Liu, A. Wang, K. Sharshov, Y. Li, Comparative metagenomics of the gut microbiota in wild greylag geese (Anser anser) and ruddy shelducks (Tadorna ferruginea), Microbiologyopen 8 (2019) e00725.
[2] L. Liu, C. Du, Y. Sun, J. Liu, Z. Pu, X. Liu, Trace element distribution in tissues and risk of exposure of ruddy shelduck wintering in Nanhaizi Wetland, Baotou, China, Environ. Sci. Pollut. Res. Int. 27 (2020) 6429–6437.
[3] J. Liu, H. Xiao, F. Lei, Q. Zhu, K. Qin, X.-W. Zhang, X.-L. Zhang, D. Zhao, G. Wang, Y. Feng, J. Ma, W. Liu, J. Wang, G.F. Gao, Highly pathogenic H5N1 influenza virus infection in migratory birds, Science 309 (2005) 1206.
[4] H. Chen, Y. Li, Z. Li, J. Shi, K. Shinya, G. Deng, Q. Qi, G. Tian, S. Fan, H. Zhao, Y. Sun, Y. Kawaoka, Properties and dissemination of H5N1 viruses isolated during an influenza outbreak in migratory waterfowl in western China, J. Virol. 80 (2006) 5976–5983.
[5] N. Komar, West Nile viral encephalitis, Rev. Sci. Tech. 19 (2000) 166–176.
[6] D.E. Swayne, J.R. Beck, C.S. Smith, W.J. Shieh, S.R. Zaki, Fatal encephalitis and myocarditis in young domestic geese (Anser anser domesticus) caused by West Nile virus, Emerg. Infect. Dis. 7 (2001) 751–753.
[7] T. Yun, D. Zhang, X. Ma, Z. Cao, L. Chen, Z. Ni, W. Ye, B. Yu, J. Hua, Y. Zhang, C. Zhang, Complete genome sequence of a novel flavivirus, duck tembusu virus, isolated from ducks and geese in China, J. Virol. 86 (2012) 3406–3407.
[8] T. Shan, S. Yang, H. Wang, H. Wang, J. Zhang, G. Gong, Y. Xiao, J. Yang, X. Wang, J. Lu, M. Zhao, Z. Yang, X. Lu, Z. Dai, Y. He, X. Chen, R. Zhou, Y. Yao, N. Kong, J. Zeng, K. Ullah, X. Wang, Q. Shen, X. Deng, J. Zhang, E. Delwart, G. Tong, W. Zhang, Virome in the cloaca of wild and breeding birds revealed a diversity of significant viruses, Microbiome 10 (2022) 60.
[9] S. Yang, D. Zhang, Z. Ji, Y. Zhang, Y. Wang, X. Chen, Y. He, X. Lu, R. Li, Y. Guo, Q. Shen, L. Ji, X. Wang, Y. Li, W. Zhang, Viral metagenomics reveals diverse viruses in tissue samples of diseased pigs, Viruses 14 (2022).
[10] M. Wille, J.-S. Eden, M. Shi, M. Klaassen, A.C. Hurt, E.C. Holmes, Virus-virus interactions and host ecology are associated with RNA virome structure in wild birds, Mol. Ecol. 27 (2018) 5263–5278.
[11] M. Wille, M. Shi, M. Klaassen, A.C. Hurt, E.C. Holmes, Virome heterogeneity and connectivity in waterfowl and shorebird communities, ISME J. 13 (2019) 2603–2616.
[12] W. Zhang, L. Li, X. Deng, B. Kapusinszky, P.A. Pesavento, E. Delwart, Faecal virome of cats in an animal shelter, J. Gen. Virol. (2014).
[13] W. Zhang, L. Li, X. Deng, J. Blümel, C.M. Nübling, A. Hunfeld, S.A. Baylis, E. Delwart, Viral nucleic acids in human plasma pools, Transfusion 56 (2016) 2248–2255.
[14] 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, Virome comparisons in wild-diseased and healthy captive giant pandas, Microbiome 5 (2017) 90.
[15] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S.M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.W. Lam, J. Wang, SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler, GigaScience (2012).
[16] A. Aiyar, The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment, Methods Mol. Biol. (2000).
[17] M. Krupovic, A. Varsani, Naryaviridae, Nenyaviridae, and Vilyaviridae: three new families of single-stranded DNA viruses in the phylum Cressdnaviricota, Arch. Virol. 167 (2022) 2907–2921.
[18] M. Sait, M. Livingstone, R. Graham, N.F. Inglis, N. Wheelhouse, D. Longbottom, Identification, sequencing and molecular analysis of Chp4, a novel chlamydiaphage of Chlamydophila abortus belonging to the family Microviridae, J. Gen. Virol. 92 (2011) 1733–1737.
[19] M.J. Tisza, C.B. Buck, A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases, Proc. Natl. Acad. Sci. U. S. A. 118 (2021).
[20] K. Rosario, A. Dayaram, M. Marinov, J. Ware, S. Kraberger, D. Stainton, M. Breitbart, A. Varsani, Diverse circular ssDNA viruses discovered in dragonflies (Odonata: epiprocta), J. Gen. Virol. 93 (2012) 2668–2681.
[21] D. Muzyka, M. Pantin-Jackwood, B. Stegniy, O. Rula, V. Bolotin, A. Stegniy, A. Gerilovych, P. Shutchenko, M. Stegniy, V. Koshelev, K. Maiorova, S. Tkachenko, N. Muzyka, L. Usova, C.L. Afonso, Wild bird surveillance for avian paramyxoviruses in the Azov-black sea region of Ukraine (2006 to 2011) reveals epidemiological connections with Europe and Africa, Appl. Environ. Microbiol. 80 (2014) 5427–5438.

[22] E. Spackman, D.E. Swayne, M. Gilbert, D.O. Joly, W.B. Karesh, D.L. Suarez, R. Sodnomdarjaa, P. Dulam, C. Cardona, Characterization of low pathogenicity avian influenza viruses isolated from wild birds in Mongolia 2005 through 2007, Virol. J. 6 (2009) 190.
[23] N.M. Nemeth, J.D. Brown, D.E. Stallknecht, E.W. Howerth, S.H. Newman, D.E. Swayne, Experimental infection of bar-headed geese (Anser indicus) and ruddy shelducks (Tadorna ferruginea) with a clade 2.3.2 H5N1 highly pathogenic avian influenza virus, Vet. Pathol. 50 (2013) 961–970.
[24] L.A. Ramírez-Martínez, E. Loza-Rubio, J. Mosqueda, M.L. González-Garay, G. García-Espinosa, Fecal virome composition of migratory wild duck species, PLoS One 13 (2018) e0206970.
[25] L. Bian, Y. Wang, X. Yao, Q. Mao, M. Xu, Z. Liang, Coxsackievirus A6: a new emerging pathogen causing hand, foot and mouth disease outbreaks worldwide, Expert Rev. Anti Infect. Ther. 13 (2015) 1061–1071.
[26] J.L. Huaman, C. Pacioni, S. Sarker, M. Doyle, D.M. Forsyth, A. Pople, T.G. Carvalho, K.J. Helbig, Novel picornavirus detected in wild deer: identification, genomic characterisation, and prevalence in Australia, Viruses 13 (2021).
[27] C.K. Yinda, R. Zell, W. Deboutte, M. Zeller, N. Conceição-Neto, E. Heylen, P. Maes, N.J. Knowles, S.M. Ghogomu, M. Van Ranst, J. Matthijnssens, Highly diverse population of Picornaviridae and other members of the Picornavirales, in Cameroonian fruit bats, BMC Genom. 18 (2017) 249.
[28] A. Philipps, M. Dauber, M. Groth, H. Schirrmeier, M. Platzer, A. Krumbholz, P. Wutzler, R. Zell, Isolation and molecular characterization of a second serotype of the encephalomyocarditis virus, Vet. Microbiol. 161 (2012) 49–57.
[29] P. Pankovics, Á. Boros, T. Kiss, G. Reuter, Identification and complete genome analysis of kobuvirus in faecal samples of European roller (Coracias garrulus): for the first time in a bird, Arch. Virol. 160 (2015) 345–351.
[30] C.M. Kinsella, A. Bart, M. Deijs, P. Broekhuizen, J. Kaczorowska, M.F. Jebbink, T. van Gool, M. Cotten, L. van der Hoek, Entamoeba and Giardia parasites implicated as hosts of CRESS viruses, Nat. Commun. 11 (2020) 4620.
[31] P.C. Kirchberger, H. Ochman, Resurrection of a global, metagenomically defined gokushovirus, Elife 9 (2020), https://doi.org/10.7554/eLife.51599.
[32] M. Walters, M. Bawuro, A. Christopher, A. Knight, S. Kraberger, D. Stainton, H. Chapman, A. Varsani, Novel single-stranded DNA virus genomes recovered from chimpanzee feces sampled from the mambilla plateau in Nigeria, Genome Announc. 5 (2017).
[33] K.L. Brentlinger, S. Hafenstein, C.R. Novak, B.A. Fane, R. Borgon, R. McKenna, M. Agbandje-McKenna, Microviridae, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus, J. Bacteriol. 184 (2002) 1089–1094.
[34] S. François, O.G. Pybus, Towards an understanding of the avian virome, J. Gen. Virol. 101 (2020) 785–790.