

Characterization of the genomic sequence of a circo-like virus and of three chaphamaparvoviruses detected in mute swan (*Cygnus olor*)

Sarah François,^{1,2} Sarah C. Hill,³ Christopher M. Perrins,¹ Oliver G. Pybus^{1,3}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the complete genomes of four ssDNA viruses: a circular replication-associated protein-encoding single-stranded DNA virus belonging to a clade previously detected only in mammals, and three chaphamaparvoviruses, which were detected by viromic surveillance of mute swan (*Cygnus olor*) fecal samples from the United Kingdom.

KEYWORDS wildlife, swan, waterbird, surveillance, viral metagenomics, virus discovery, parvovirus, CRESS DNA virus

Our knowledge of viruses infecting wild birds remains scarce, which is detrimental to poultry health and wildlife conservation (1, 2).

We processed seven mute swan (*Cygnus olor*) non-invasive samples collected in United Kingdom between 2016 and 2019 [for details, see reference (3)]. About 0.5 mL of feces was collected into a tube containing 1 mL of Universal Transport Media. Tubes were shaken and kept on ice in the field, and stored at -80°C .

Viromes were obtained as described in reference (4). We followed manufacturers' instructions and default parameters except where otherwise noted. Samples were homogenized by a bead beater, filtered through a $0.45\ \mu\text{m}$ filter, digested by DNaseI and RNaseA incubation at 37°C for 1.5 h. DNA and RNA were extracted using a QIAamp Viral RNA Mini Kit. Reverse transcription was performed using a SuperScript IV VILO kit, cDNAs were purified by a QIAquick PCR Purification Kit, and dsDNA was synthesised by Klenow DNA polymerase I. DNA was amplified by random PCR amplification (Q5 Hot Start High-Fidelity kit). PCR products were purified using a NucleoSpin gel and PCR clean-up kit. Libraries were prepared using a NEB NEXT Ultra II DNA Library prep kit, and sequenced on a NovaSeq6000 in 2×150 bp paired-end mode.

Adaptors were removed and reads were filtered for quality (q30 and length >45 nt) using cutadapt 2.19 (5), and 153,109,590 paired-end reads were assembled into contigs by MEGAHIT 1.2.9 (6). Taxonomic assignment was achieved using DIAMOND 0.9.30 against the NCBI nr protein database (7). Genome coverage was assessed by mapping using Bowtie2 3.5.1 (local sensitive) (8). Open reading frames (ORFs) were identified using ORF finder (length cutoff >300 nt) on Geneious Prime 2022.0.2 (9), and were annotated by blastp query-centered alignment against RefSeq viral database on 18 September 2023.

We reconstructed the complete circular genome of mute swan circo-like virus (MSCLV; length: 3,663 nt; GC content: 35.6%; average coverage depth: 298; 9,968 mapped reads, [SRR26091305](#)) and confirmed it through Sanger sequencing of PCR amplicons using GoTaQ HotStar kit with overlapping primers. Chromatograms were checked for disparities. MSCLV genome contained a replication-associated protein gene (918 nt – predicted amino acid sequence: 306 aa), a capsid protein gene (507 nt –

Editor Jelle Matthijnssens, Katholieke Universiteit Leuven, Belgium

Address correspondence to Sarah François, sarah.francois@inrae.fr.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 12 December 2023

Accepted 7 February 2024

Published 20 February 2024

Copyright © 2024 François et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

TABLE 1 Information on the three chaphamaparvoviruses reconstructed from mute swan viromic data

| Virus | Genome | | Coverage | | | Putative proteins | | | Closest identified relatives | | | |
|-----------------------------------|-----------|------|----------|--------|-------------|-------------------|-----------|-----------|--|------------------|----------------------|--------------------------|
| | Size (nt) | %GC | Average | Number | Sample | Name | Size (nt) | Size (AA) | Virus name | Accession number | AA pairwise identity | Host name |
| Chaphamaparvovirus anseriform7 | 4,370 | 41.9 | 50 | 2,019 | SRR26091311 | NS1 | 2,007 | 669 | <i>Aegithalos caudatus parvoviridae</i> sp. | QTE03727 | 79.60% | <i>Cygnus atratus</i> |
| | | | | | | NS2 | 594 | 198 | Wood duck chaphamaparvovirus | QMI57945 | 73.20% | <i>Chenonetta jubata</i> |
| | | | | | | NS3 | 438 | 146 | Chestnut teal chaphamaparvovirus 1 | YP_010802862 | 68.30% | <i>Anas castanea</i> |
| | | | | | | VP | 1,671 | 557 | <i>Cygnus atratus</i> Chaphamaparvovirus | QTE04016 | 61.90% | <i>Cygnus atratus</i> |
| Chaphamaparvovirus anseriform8 | 4,296 | 39.5 | 230 | 9,206 | SRR26091304 | NS1 | 2,052 | 684 | <i>Parvoviridae</i> sp. | QKE54873 | 50.50% | Unspecified bird |
| | | | | | | NS2 | 621 | 207 | Chestnut teal chaphamaparvovirus | QMI57883 | 50.50% | <i>Anas castanea</i> |
| | | | | | | NS3 | 429 | 143 | Chestnut teal chaphamaparvovirus | QMI57870 | 49.30% | <i>Anas castanea</i> |
| | | | | | | VP | 1,626 | 542 | <i>Parvoviridae</i> sp. | QKE54874 | 45.90% | Unspecified bird |
| Chaphamaparvovirus anseriform9 | 4,432 | 39.5 | 206 | 8,343 | SRR26091311 | NS1 | 2,007 | 669 | Mute swan feces-associated chapparvovirus 6 | QUS52585 | 72.60% | <i>Cygnus olor</i> |
| | | | | | | NS2 | 606 | 202 | Chestnut teal chaphamaparvovirus 1 | QMI57883 | 62.10% | <i>Anas castanea</i> |
| | | | | | | NS3 | 447 | 149 | Chestnut teal chaphamaparvovirus 1 | YP_010802862 | 66.00% | <i>Anas castanea</i> |
| | | | | | | VP | 1,689 | 563 | Mute swan feces-associated chapparvovirus 6 | QUS52584 | 69.90% | <i>Cygnus olor</i> |

169 aa), and a putative origin of replication marked by a conserved nonamer motif (TACTAAAGTA) flanked by a stem-loop structure (10). The closest relatives of MSCLV are pig-infecting circo-like viruses (11) [Po-Circo-like virus isolate CZH12 (MW881210) with which MSCLV shared 50.8% replication-associated protein pairwise identity; and Po-Circo-like virus HN39-01 (OP302752), 28.4% capsid protein identity] (Fig. 1). Based on the most conserved species demarcation threshold for circular replication-associated protein-encoding single-stranded DNA virus families (i.e., 77% genome-wide identity), MSCLV putatively belongs to a divergent species (12).

We report the complete CDS (coding sequence) of three members of the mammal and bird infecting *Chaphamaparvovirus* genus (*Parvoviridae* family, *Hamaparvovirinae* subfamily, 10.6084 /m9.figshare.24777786). Their closest relatives are bird-associated chaphamaparvoviruses from wild Anatidae samples, with which they shared between 50.5% and 79.6% non-structural protein 1 (NS1) protein identity (Table 1). Based on the *Parvoviridae* family species demarcation threshold (i.e., 85% NS1 protein identity), these viruses could belong to novel species (13).

ACKNOWLEDGMENTS

S.F. and O.G.P. were supported by BBSRC grant BB/T008806/1. S.F. was supported by Roslin Institute career grant from the UK International Coronavirus Network (UK-ICN). S.C.H. was supported by the Wellcome Trust (102427/Z/13/Z and 220414/Z/20/Z).

We thank Mrs C. Townsend for permission to study the swans.

AUTHOR AFFILIATIONS

¹Department of Biology, University of Oxford, Oxford, United Kingdom

²DGIMI, Univ Montpellier, INRAE, Montpellier, France

³Department of Pathobiology and Population Science, Royal Veterinary College, Hatfield, United Kingdom

AUTHOR ORCID*s*

Sarah François  <http://orcid.org/0000-0002-1138-8800>

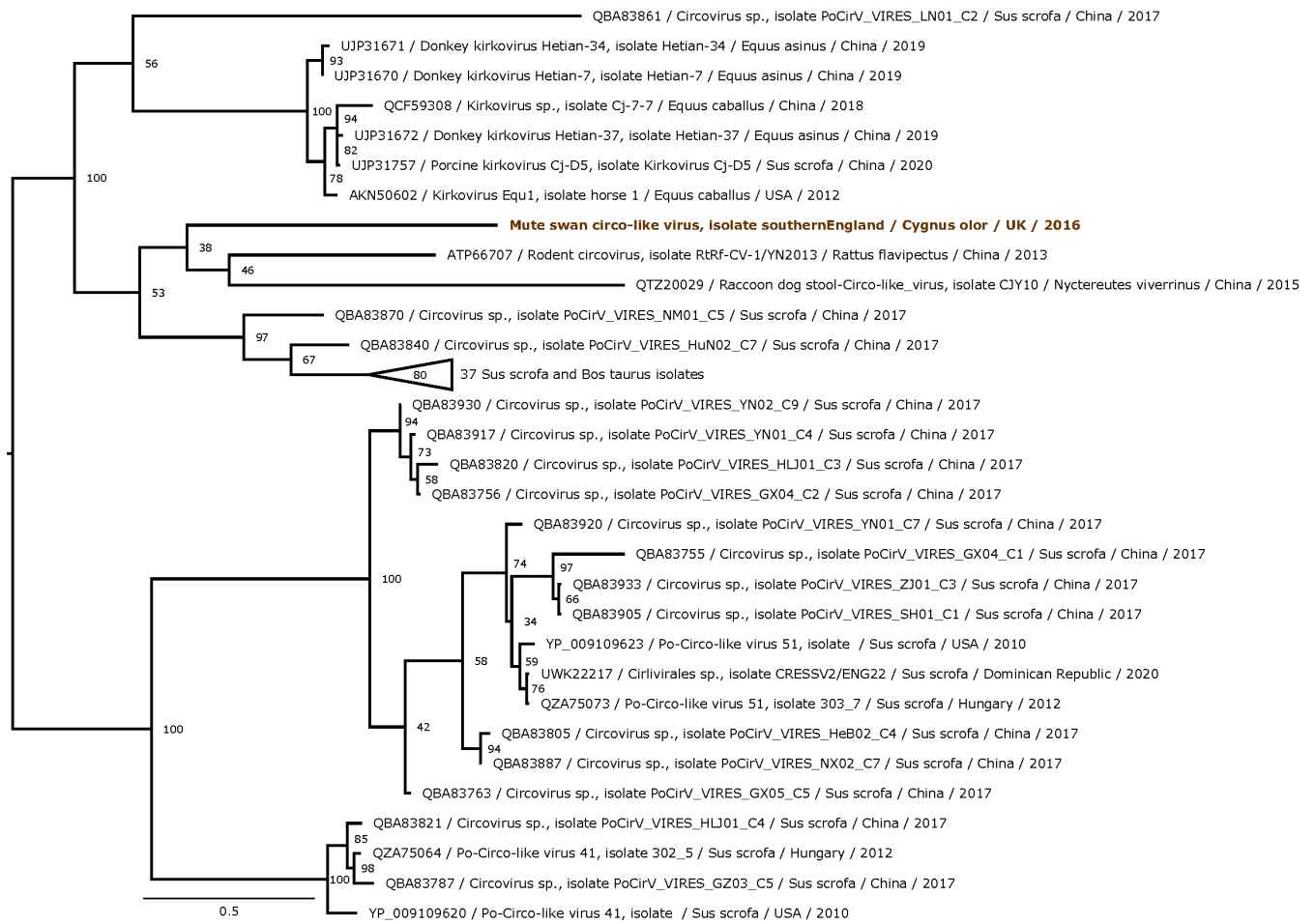


FIG 1 Maximum likelihood phylogenetic tree based on the capsid protein of the MSCLV and its 65 closest relatives. Protein sequences used in phylogenetic analyses were obtained by blastx from the NCBI nr database (18 September 2023). Proteins were aligned using MAFFT 7.450 with the L-INS-i algorithm. Maximum likelihood trees were estimated using RAxML 8.2.11, under the LG + G + I + F protein evolution model. Branch support was evaluated using 100 bootstrapped replicates. Trees were mid-point rooted and visualized with MEGAX 10.2.6. Bootstrap values (100 replicates) >30% are indicated at each node. The scale bar corresponds to expected amino acid substitutions per site. The sequence obtained from our sample is in bold red.

FUNDING

| Funder | Grant(s) | Author(s) |
|---|-------------------------------|-----------------------------------|
| UKRI Biotechnology and Biological Sciences Research Council (BBSRC) | BB/T008806/1 | Sarah François Oliver G. Pybus |
| UK International Coronavirus Network | Roslin Institute career grant | Sarah François |
| Wellcome Trust (WT) | 102427/Z/13/Z, 220414/Z/20/Z | Sarah C. Hill |

AUTHOR CONTRIBUTIONS

Sarah François, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Sarah C. Hill, Resources, Writing – review and editing | Christopher M. Perrins, Methodology, Writing – review and editing | Oliver G. Pybus, Conceptualization, Funding acquisition, Writing – review and editing, Supervision

DATA AVAILABILITY

The genomic sequences of mute swan circo-like virus (MSCLV), Chaphamaparvovirus anseriform7, Chaphamaparvovirus anseriform8, and Chaphamaparvovirus anseriform9 have been deposited at GenBank under the accession numbers [OR583913](#), [OR583914](#), [OR583915](#), and [OR583916](#). High-throughput sequencing reads and raw Sanger reads were deposited in SRA under the accession no. [SRR26091304](#) to [SRR26091311](#) and [SRR27606811](#) under [PRJNA685791](#) BioProject.

REFERENCES

1. Olsen B, Munster VJ, Wallensten A, Waldenström J, Osterhaus ADME, Fouchier RAM. 2006. Global patterns of influenza A virus in wild birds. *Science* 312:384–388. <https://doi.org/10.1126/science.1122438>
2. François S, Pybus OG. 2020. Towards an understanding of the avian virome. *J Gen Virol* 101:785–790. <https://doi.org/10.1099/jgv.0.001447>
3. Hill SC, François S, Thézé J, Smith AL, Simmonds P, Perrins CM, van der Hoek L, Pybus OG. 2022. Impact of host age on viral and bacterial communities in a waterbird population. *ISME J* 17:215–226. <https://doi.org/10.1038/s41396-022-01334-4>
4. François S, Filloux D, Fernandez E, Ogliaastro M, Roumagnac P. 2018. Viral metagenomics approaches for high-resolution screening of multiplexed arthropod and plant viral communities. *Methods Mol Biol* 1746:77–95. https://doi.org/10.1007/978-1-4939-7683-6_7
5. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 17:10. <https://doi.org/10.14806/ej.17.1.200>
6. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
7. Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
9. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
10. Zhao L, Rosario K, Breitbart M, Duffy S. 2019. Eukaryotic circular replicating single-stranded DNA (CRESS DNA) viruses: ubiquitous viruses with small genomes and a diverse host range. *Adv Virus Res* 103:71–133. <https://doi.org/10.1016/bs.aivir.2018.10.001>
11. Yang K, Zhang M, Liu Q, Cao Y, Zhang W, Liang Y, Song X, Ji K, Shao Y, Qi K, Tu J. 2021. Epidemiology and evolution of emerging porcine circovirus-like viruses in pigs with hemorrhagic dysentery and diarrhea symptoms in central China from 2018 to 2021. *Viruses* 13:2282. <https://doi.org/10.3390/v13112282>
12. Rosario K, Mettel KA, Benner BE, Johnson R, Scott C, Yusseff-Vanegas SZ, Baker CCM, Cassill DL, Storer C, Varsani A, Breitbart M. 2018. Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates. *PeerJ* 6:e5761. <https://doi.org/10.7717/peerj.5761>
13. Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger A-M, Hughes J, Mietzsch M, Modha S, Ogliaastro M, Pénzes JJ, Pintel DJ, Qiu J, Soderlund-Venermo M, Tattersall P, Tijssen P. 2019. ICTV virus taxonomy profile: parvoviridae. *J Gen Virol* 100:367–368. <https://doi.org/10.1099/jgv.0.001212>