Complete Genome Sequence of a Phapecoctavirus Isolated from a Pigeon Cloacal Swab Sample

Anthony Khalifeh,[a] Simona Kraberger,[a] Daria Dziewulska,[b] Tomasz Stenzel,[b] Arvind Varsani[a,c]

[a]Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[b]Department of Poultry Diseases, Faculty of Veterinary Medicine, University of Warmia and Mazury, Olsztyn, Poland
[c]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town, South Africa

ABSTRACT The complete genome sequence of a bacteriophage in the genus *Phapecoctavirus* (family *Myoviridae*) isolated from a cloacal swab specimen from a domestic pigeon (*Columba livia* f. *domestica*) was identified using a high-throughput sequencing approach. The genome is 150,892 bp with a GC content of 39.1%, containing 269 open reading frames and 11 tRNA genes.

Over the past decade, with the use of high-throughput sequencing, a plethora of known and novel viral sequences have been identified from a variety of sample types. For avian samples, cloacal swab specimens have been used for the identification of various pathogenic viruses, such as circoviruses, flaviviruses, gyroviruses, orthomyxoviruses, papillomaviruses, paramyxoviruses, and polyomaviruses. Additionally, numerous bacteriophages have been identified, including those classified in the *Myoviridae* family (1). Myoviruses are double-stranded DNA (dsDNA) viruses that have contractile tails (2). Identifying new viral genomes helps provide insights into viral diversity, evolution, and putative hosts in the case of bacteriophages. Here, we describe a new member of the *Phapecoctavirus* genus (family *Myoviridae*) that was identified in a cloacal swab specimen collected from an 8-week-old racing/carrier domestic pigeon (*Columba livia* f. *domestica*) from an amateur/hobby pigeon facility located in northern Poland (Samolubie, Bartoszyckie County, Warmińskpo-Mazurskie voivodeship; GPS coordinates, 54°11′02.1″N, 20°43′45.2″E).

Viral DNA was extracted from a pigeon cloacal swab specimen using the High Pure viral nucleic acid kit (Roche Diagnostics, USA). Viral DNA was amplified using rolling circular amplification (RCA) with the TempliPhi 2000 kit (GE Healthcare, USA). The resulting RCA DNA was used to generate 2 × 150-bp libraries at BGI (Hong Kong) using their DNBseq normal DNA library option, and the libraries were sequenced on their BGIseq sequencer. The resulting reads (11,610,789 read pairs) were quality trimmed using Trimmomatic v0.39 (3), and the trimmed reads were *de novo* assembled using metaSPAdes v3.12.0 (4). Contigs of >1,000 nucleotides (nt) were analyzed against an NCBI RefSeq (5) viral protein sequence database, and bacteriophages were identified using VirSorter (6). All tools were run with default parameters unless otherwise specified. A circular contig (based on terminal redundancy) of 150,892 nt (GC content, 39.1%) was identified that is most closely related to viruses in the genus *Phapecoctavirus* (family *Myoviridae*). A total of 36,072 reads mapped to this bacteriophage genome with a mean coverage of 16×. RASTtk (7) was used to annotate this genome and predicted 269 open reading frames (varying in size from 96 to 3,333 nt) and 11 tRNA genes (Arg, Asn, Gln, Gly, Ile, Met, Met, Pro, Ser, Thr, and Tyr). We tentatively name this virus dompiswa virus (domestic pigeon swab-associated virus).

All related genomes belonging to the genus *Phapecoctavirus* (*n* = 15) were downloaded from GenBank (on 12 November 2020). The genomes were linearized at the
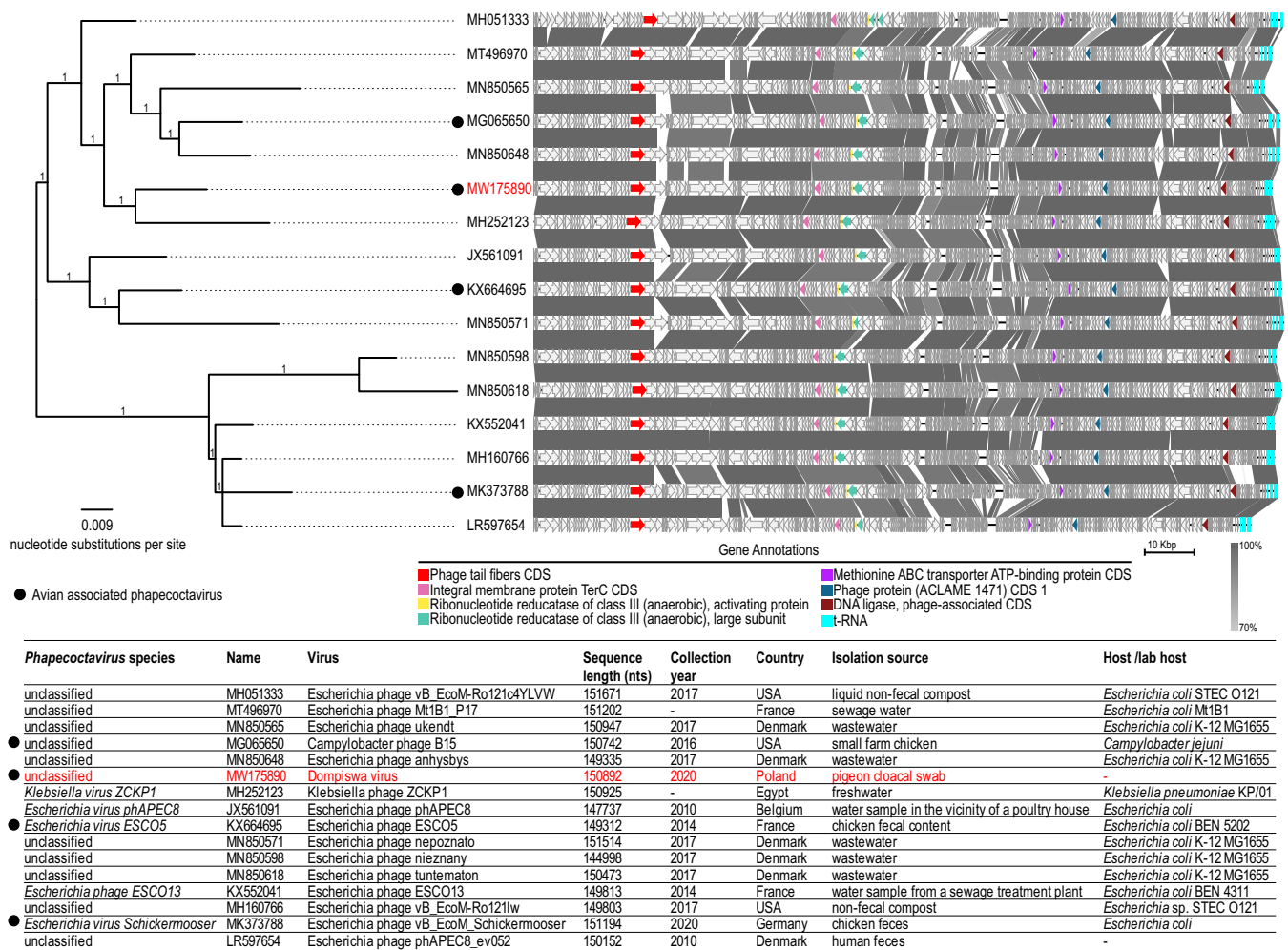
FIG 1 Phylogenetic analysis of phapecoctavirus genomes (*n* = 16). The accession number of the dompiswa virus and associated information are highlighted in red. Comparisons of phapecoctavirus genomes are shown to the right of the phylogeny. Gene annotations are highlighted to show relative gene placement along the genomes. Gray boxes between sequences indicate similarity of regions based on BLASTn analysis. CDS, coding sequence.

Phapecoctavirus species	Name	Virus	Sequence length (nts)	Collection year	Country	Isolation source	Host /lab host
unclassified	MH051333	Escherichia phage vB_EcoM-Ro121c4YLVW	151671	2017	USA	liquid non-fecal compost	*Escherichia coli* STEC O121
unclassified	MT496970	Escherichia phage Mt1B1_P17	151202	-	France	sewage water	*Escherichia coli* Mt1B1
unclassified	MN850565	Escherichia phage ukendt	150947	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
● unclassified	MG065650	Campylobacter phage B15	150742	2016	USA	small farm chicken	*Campylobacter jejuni*
unclassified	MN850648	Escherichia phage anhysbys	149335	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
● unclassified	MW175890	Dompiswa virus	150892	2020	Poland	pigeon cloacal swab	-
Klebsiella virus ZCKP1	MH252123	Klebsiella phage ZCKP1	150925	-	Egypt	freshwater	*Klebsiella pneumoniae* KP/01
Escherichia virus phAPEC8	JX561091	Escherichia phage phAPEC8	147737	2010	Belgium	water sample in the vicinity of a poultry house	*Escherichia coli*
● *Escherichia virus ESCO5*	KX664695	Escherichia phage ESCO5	149312	2014	France	chicken fecal content	*Escherichia coli* BEN 5202
unclassified	MN850571	Escherichia phage nepoznato	151514	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
unclassified	MN850598	Escherichia phage nieznany	144998	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
unclassified	MN850618	Escherichia phage tuntematon	150473	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
Escherichia phage ESCO13	KX552041	Escherichia phage ESCO13	149813	2014	France	water sample from a sewage treatment plant	*Escherichia coli* BEN 4311
unclassified	MH160766	Escherichia phage vB_EcoM-Ro121lw	149803	2017	USA	non-fecal compost	*Escherichia* sp. STEC O121
● *Escherichia virus Schickermooser*	MK373788	Escherichia phage vB_EcoM_Schickermooser	151194	2020	Germany	chicken feces	*Escherichia coli*
unclassified	LR597654	Escherichia phage phAPEC8_ev052	150152	2010	Denmark	human feces	-

end of the cluster of tRNA genes and aligned with MAFFT (8), and the resulting alignment was used to infer a maximum likelihood phylogenetic tree with PhyML (9) using the WAG+G+I nucleotide substitution model (determined to be the best-fit model using ModelTest [10]). Branches with approximate likelihood ratio test support of <0.8 were collapsed using TreeGraph 2 (11) and midpoint rooted. Phylogenetically, dompiswa virus is most closely related to *Klebsiella* phage ZCKP1 (GenBank accession number MH252123), isolated from freshwater (12), sharing 92.3% intergenomic distance determined using VIRIDIC (13), and it clusters with other unclassified phages (GenBank accession numbers MH051333, MT496970, MN850565, MG065650, and MN850648) (Fig. 1).

Five of the 16 phapecoctaviruses have been identified from avian fecal sources, and according to culture-based laboratory approaches, 14 have been shown to infect enterobacteria (*Campylobacter jejuni*, *Escherichia coli*, or *Klebsiella pneumoniae*). Therefore, it is likely that dompiswa virus infects enterobacteria, but this needs to be confirmed.

Data availability. This genome sequence has been deposited in GenBank under the accession number MW175890, and mapped short reads have been deposited in the SRA under the accession number SRR12914778.

REFERENCES

1. Vibin J, Chamings A, Collier F, Klaassen M, Nelson TM, Alexandersen S. 2018. Metagenomics detection and characterisation of viruses in faecal samples from Australian wild birds. Sci Rep 8:8686. https://doi.org/10.1038/s41598-018-26851-1.
2. Lavigne R, Ceyssens P-J. 2012. Myoviridae, p 46–62. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy. Elsevier, San Diego, CA. https://doi.org/10.1016/b978-0-12-384684-6.00002-1.
3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
5. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. Nucleic Acids Res 43:D571–D577. https://doi.org/10.1093/nar/gku1207.
6. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.
7. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, III, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365.
8. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
9. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.
10. Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818. https://doi.org/10.1093/bioinformatics/14.9.817.
11. Stöver BC, Müller KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7. https://doi.org/10.1186/1471-2105-11-7.
12. Taha OA, Connerton PL, Connerton IF, El-Shibiny A. 2018. Bacteriophage ZCKP1: a potential treatment for *Klebsiella pneumoniae* isolated from diabetic foot patients. Front Microbiol 9:2127. https://doi.org/10.3389/fmicb.2018.02127.
13. Moraru C, Varsani A, Kropinski AM. 2020. VIRIDIC: a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. Viruses 12:1268. https://doi.org/10.3390/v12111268.