



Genome Sequence of Feline Papillomavirus Strain P20 Assembled from Metagenomic Data from the Skin of a House Cat Owner

Emma H. Graham,^a Michael S. Adamowicz,^b Peter C. Angeletti,^{c,d} Jennifer L. Clarke,^{e,f} Samodha C. Fernando,^g
Joshua R. Herr^{d,h,i}

^aPh.D. Program in Complex Biosystems, University of Nebraska, Lincoln, Nebraska, USA

^bCollege of Agricultural Sciences and Natural Resources, University of Nebraska, Lincoln, Nebraska, USA

^cNebraska Center for Virology, University of Nebraska, Lincoln, Nebraska, USA

^dSchool of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA

^eDepartment of Statistics, University of Nebraska, Lincoln, Nebraska, USA

^fFood Science and Technology Department, University of Nebraska, Lincoln, Nebraska, USA

^gDepartment of Animal Science, University of Nebraska, Lincoln, Nebraska, USA

^hDepartment of Plant Pathology, University of Nebraska, Lincoln, Nebraska, USA

ⁱCenter for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska, USA

ABSTRACT A feline papillomavirus genome was assembled from metagenomic sequencing data collected from the skin of a house cat owner. The circular genome of strain P20 is 8,069 bp in length, has a GC content of 54.38%, and displays genome organization typical of feline papillomaviruses. The genome exhibits approximately 75% sequence similarity to other feline papillomavirus genomes.

Papillomaviruses broadly infect mammals, reptiles, and birds (1). Conservation of the major capsid protein L1 suggests that the evolution of papillomaviruses mirrors the phylogeny of their hosts (2). Recent studies have reported on the discovery and diversity of papillomaviruses associated with domestic and wild cats (3). Seven types of feline papillomaviruses, which may or may not produce both skin and oral squamous cell carcinomas in house cats, are currently recognized (4).

We previously studied the diversity and stability of DNA viruses on human skin by sampling three anatomical locations, i.e., left hand, right hand, and scalp, longitudinally over a 6-month period for 43 human subjects (5). Briefly, samples were collected using a tandem dry and wet swab technique. Swabs were then saturated in 1 × phosphate-buffered saline (PBS) and centrifuged at 20,000 × *g*. The sample eluent was run through a 0.22- μ m filter to remove cellular and bacterial contaminants, and the resulting filtrate was used for viral DNA extraction using the QIAmp ultrasensitive virus kit (Qiagen, Hilden, Germany) and whole-genome amplification (WGA) using the TruePrime WGA kit (Syngen Biotechnology, Inc., Taipei City, Taiwan), each according to the manufacturer's protocol. The resulting DNA was sheared to 600 bp prior to library preparation using the NEBNext Ultra II library preparation kit (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol. The libraries were then sequenced as 150-bp paired-end reads on the HiSeq 2500 platform (Illumina, Inc., San Diego, CA, USA). We identified one participant with a metagenome that consistently showed the presence of a feline papillomavirus, who self-identified as being an owner of a domesticated house cat (5).

To investigate further, 15 metagenome samples from that specific human individual were mapped to a *Papillomaviridae* reference database compiled from NCBI using BBMap v38.94, with a minimum match of 95% base pair similarity and a *k* value of 13 (6). Mapped reads were processed using khmer v2.0.0 (7) to remove singletons and overly abundant reads.

Editor Jelle Matthijnsens, KU Leuven

Copyright © 2022 Graham et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Joshua R. Herr, joshua.r.herr@gmail.com.

The authors declare no conflict of interest.

Received 28 December 2021

Accepted 7 June 2022

Published 5 July 2022

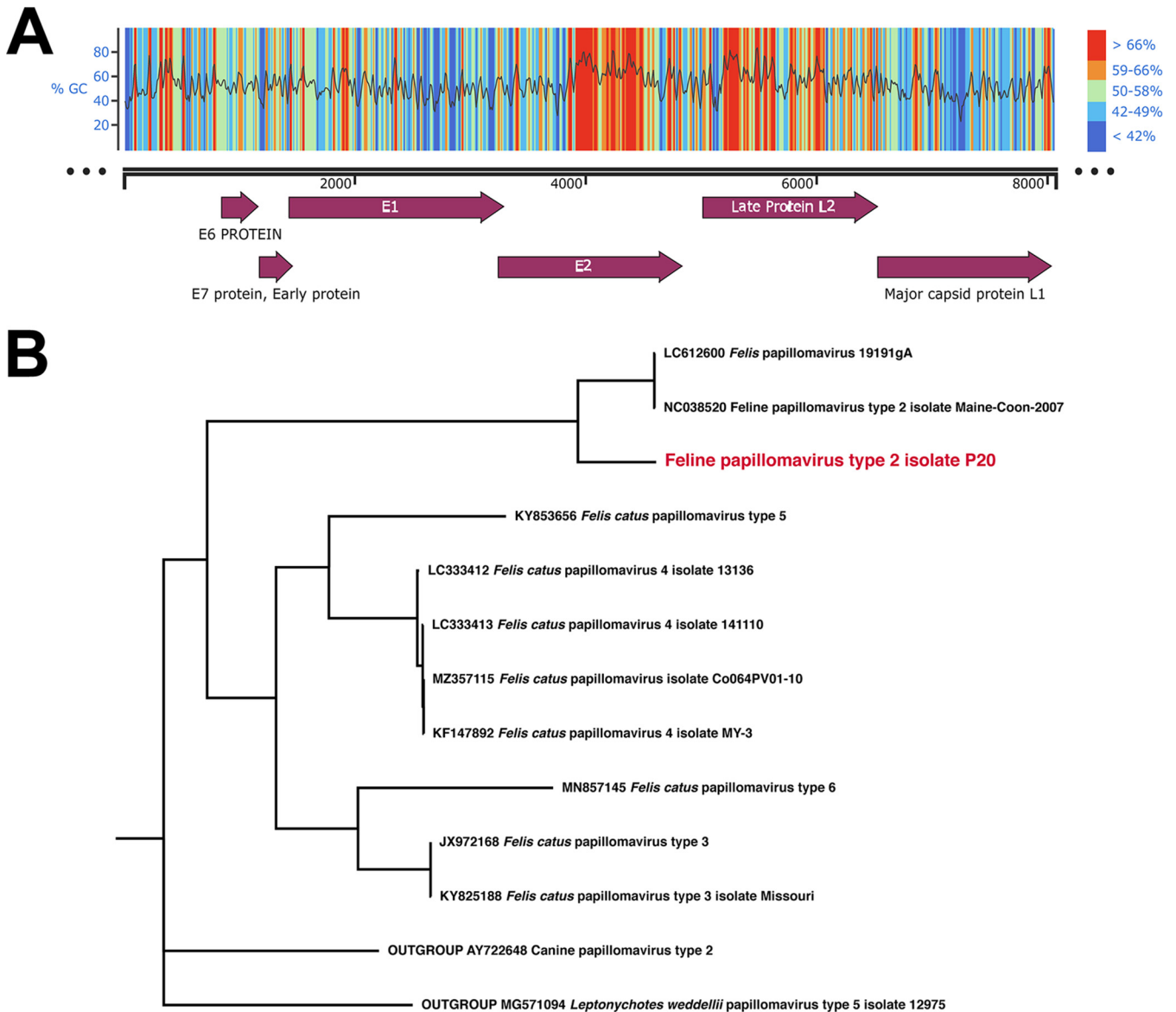


FIG 1 (A) Linear genome map (8,069 bp in length) of the feline papillomavirus type 2-like virus. The top portion shows the GC contents for regions of the circular genome. Higher GC contents are shown in warmer colors (red), and lower GC contents are shown in cooler colors (blue). The locations of the six gene regions in the genome are shown below, with open reading frame annotations (purple arrows) for the E6 protein, E7 protein, E1 protein, E2 protein, late protein L2, and major capsid protein L1. (B) Phylogenetic tree of the placement of the feline type 2 papillomavirus identified in this study along with nine other feline papillomaviruses. Identifiers on the phylogenetic tree correspond to NCBI Genbank accession numbers. The tree was rooted with canine papillomavirus type 2 and *Leptonychotes weddellii* (Weddell seal) papilloma virus type 5.

Filtered papillomavirus reads were *de novo* assembled using MEGAHIT v1.2.8 (8), and assembly quality was assessed using QUAST v5.0.2 (9). Read coverage for the assembly was an average of 295.18 reads per base, with a standard deviation of 46.61 reads per base. A single contig was evaluated for completeness using CheckV v0.7.0 (10) and nucleotide-based classification tools, such as Kraken2 v2.0.8-beta (11), Demovir (<https://github.com/feargalr/Demovir>), and BLASTn (with a >10% query coverage cutoff value) (12), as well as the protein-coding classification tool Kaiju v1.7 (13). All tools were run with default parameters unless otherwise specified.

We assembled a complete circular feline papillomavirus genome, identified as strain P20, that was 8,069 bp in length and exhibited a GC content of 54.38%. Six open reading frames, including the E6 protein, E7 protein, E1 protein, E2 protein, late protein L2, and major capsid protein L1, which are shared among many animal papillomaviruses (1), were annotated with

Prokka v1.14.5 (14) using the embedded viral annotation database and Cenote-Taker 2 v2.1 (15) (Fig. 1) (visualized with SnapGene; GSL Biotech LLC, San Diego, CA). Using BLASTn, our genome assembly showed the greatest similarity (with an average of 70.3% shared amino acids) to feline papillomavirus type 2 (family *Papillomaviridae*, genus *Dyothetapapillomavirus*) isolated from the skin of a domestic Maine Coon house cat in 2007 (GenBank accession number [NC_038520](https://www.ncbi.nlm.nih.gov/nuccore/NC_038520)).

To determine the phylogenetic placement of strain P20, we downloaded all feline papillomavirus genomes available in the NCBI database (along with two outgroups from closely related hosts, namely, domestic dog and Weddell seal) and aligned the six shared annotated genes using MUSCLE v3.8.1551 (16). After removal of uninformative amino acid regions using Gblocks v0.91b (17), a phylogenetic tree (Fig. 1) was generated using IQ-TREE v1.6.12 (18) with the LG+F+I+G4 substitution model, which was determined to be optimal by ModelFinder v1.5.4 (19). This phylogeny, visualized with FigTree v1.4.4 (<https://github.com/rambaut/figtree>), placed our genome in proximity to other feline papilloma type 2 strains.

Data availability. All raw sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number [PRJNA754140](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA754140). Within that BioProject, the 15 sequencing files used in this study had the following accession numbers: [SRS9770471](https://www.ncbi.nlm.nih.gov/sra/SRS9770471), [SRS9770472](https://www.ncbi.nlm.nih.gov/sra/SRS9770472), [SRS9770476](https://www.ncbi.nlm.nih.gov/sra/SRS9770476), [SRS9770495](https://www.ncbi.nlm.nih.gov/sra/SRS9770495), [SRS9770500](https://www.ncbi.nlm.nih.gov/sra/SRS9770500), [SRS9770501](https://www.ncbi.nlm.nih.gov/sra/SRS9770501), [SRS9770587](https://www.ncbi.nlm.nih.gov/sra/SRS9770587), [SRS9770588](https://www.ncbi.nlm.nih.gov/sra/SRS9770588), [SRS9770589](https://www.ncbi.nlm.nih.gov/sra/SRS9770589), [SRS9770614](https://www.ncbi.nlm.nih.gov/sra/SRS9770614), [SRS9770616](https://www.ncbi.nlm.nih.gov/sra/SRS9770616), [SRS9770617](https://www.ncbi.nlm.nih.gov/sra/SRS9770617), [SRS9770642](https://www.ncbi.nlm.nih.gov/sra/SRS9770642), [SRS9770643](https://www.ncbi.nlm.nih.gov/sra/SRS9770643), and [SRS9770644](https://www.ncbi.nlm.nih.gov/sra/SRS9770644). The complete feline papillomavirus P20 genome has been submitted to NCBI GenBank with the accession number [OL310516](https://www.ncbi.nlm.nih.gov/nuccore/OL310516). All metadata, sequences, annotation files, and scripts used here are publicly available and archived at https://github.com/HerrLab/Graham_2021_feline_papilloma_P20.

ACKNOWLEDGMENTS

This work was completed using the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative. This work was supported by the Department of Justice (grants 2017-IJ-CX-0025, 2019-75-CX-0075, and 2019-R2-CX-0048). This funding agency had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

We thank all of the anonymous participants for their contributions to this study. We also thank members of the Fernando laboratory who assisted with advice on laboratory methods.

We have no competing interests to declare.

REFERENCES

- Rector A, Van Ranst M. 2013. Animal papillomaviruses. *Virology* 445:213–223. <https://doi.org/10.1016/j.virol.2013.05.007>.
- Rector A, Lemey P, Tachezy R, Mostmans S, Ghim S-J, Van Doorslaer K, Roelke M, Bush M, Montali RJ, Joslin J, Burk RD, Jenson AB, Sundberg JP, Shapiro B, Van Ranst M. 2007. Ancient papillomavirus-host co-speciation in Felidae. *Genome Biol* 8:R57. <https://doi.org/10.1186/gb-2007-8-4-r57>.
- Munday JS, Thomson NA. 2021. Papillomaviruses in domestic cats. *Viruses* 13:1664. <https://doi.org/10.3390/v13081664>.
- Kaynarcalidan O, Oğuzoğlu TÇ. 2021. The oncogenic pathways of papillomaviruses. *Vet Comp Oncol* 19:7–16. <https://doi.org/10.1111/vco.12659>.
- Graham EH, Fernando S, Clarke J, Herr JR, Adamowicz M. 2022. The application of the skin virome for human identification. *Forensic Sci Int Genet* 57:102662. <https://doi.org/10.1016/j.fsigen.2022.102662>.
- Bushnell B. 2014. BMAP: a fast, accurate, splice-aware aligner. <https://sourceforge.net/projects/bbmap>.
- Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edverson G, Fay S, Fenton J, Fenzl T, Fish J, Garcia-Gutierrez L, Garland P, Gluck J, González I, Guermond S, Guo J, Gupta A, Herr JR, Howe A, Hyer A, Härpfer A, Irber L, Kidd R, Lin D, Lippi J, Mansour T, McA'Nulty P, McDonald E, Mizzi J, Murray KD, Nahum JR, Nanlohy K, Nederbragt AJ, Ortiz-Zuazaga H, Ory J, Pell J, Pepe-Rannev C, Russ ZN, Schwarz E, Scott C, Seaman J, Sievert S, Simpson J, Skennerton CT, Spencer J, Srinivasan R, Standage D, et al. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res* 4:900. <https://doi.org/10.12688/f1000research.6924.1>.
- 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>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosch E, Roux S, Kyrpidis NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 39:578–585. <https://doi.org/10.1038/s41587-020-00774-7>.
- Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257–11259. <https://doi.org/10.1038/ncomms11257>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Tisza MJ, Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. 2021. Cenote-Taker

- 2 democratizes virus discovery and sequence annotation. *Virus Evol* 7:veaa100. <https://doi.org/10.1093/ve/veaa100>.
16. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
 17. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>.
 18. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
 19. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.