bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Feline Papillomavirus Strain P20 Assembled from Metagenomic Data Isolated from the Human Skin of a House Cat Owner

Ema H. Graham [1], Michael S. Adamowicz [2], Peter C Angeletti [3,4], Jennifer L. Clarke [5,6], Samodha C. Fernando [7], Joshua R. Herr [4,8,9,#]

[1] PhD Program in Complex Biosystems, University of Nebraska, 68583, Lincoln, NE, USA
[2] College of Agricultural Sciences & Natural Resources, University of Nebraska, Lincoln, NE 68583, USA
[3] Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68583
[4] School of Biological Sciences, University of Nebraska, Lincoln, NE 68588
[5] Department of Statistics, University of Nebraska, Lincoln, NE, 68588, USA
[6] Food Science and Technology Department, University of Nebraska, Lincoln, NE 68588 USA
[7] Department of Animal Science, University of Nebraska, Lincoln, NE, 68583, USA
[8] Department of Plant Pathology, University of Nebraska, Lincoln, NE 68503, USA
[9] Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68503, USA

[#] Corresponding Author
Joshua R. Herr, PhD;
422 Plant Sciences Hall
University of Nebraska
Lincoln NE 68583
Email: jherr@unl.edu
Phone: +1 (402) 472-1352

RUNNING TITLE: Feline papillomavirus from human skin metagenome

KEYWORDS
Virome, Metagenome, Human Skin, Papilloma Virus, Feline *Papillomaviridae*

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

A feline papillomavirus was assembled from metagenomic sequencing data collected from the human skin of a house cat owner. This circular papillomavirus strain P20 is 8069 bp in length, has a GC content of 54.38%, and displays genome organization typical of feline papillomaviruses with six annotated protein coding regions. The genome exhibits approximately 75% sequence similarity to other feline papillomavirus genomes.

ANNOUNCEMENT

Papillomaviruses are most recognized for infecting humans, but they also broadly infect mammals, reptiles, and birds (1). Classification of papillomaviruses has largely focused on the sequence divergence of the Major Capsid Protein L1. High conservation of the L1 protein across papillomaviruses suggests their evolution 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 are currently recognized that may or may not produce both skin and oral squamous cell carcinomas in house cats (4).

In a previous study we addressed the diversity and stability of DNA viruses on human skin by sampling three anatomical locations – left hand, right hand, and scalp – over a longitudinal 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 1x PBS and were centrifuged at 20,000xg. Sample eluent was then run through a 0.22 µm filter to remove cellular and bacterial contaminants and the resulting filtrate was used for viral

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA extraction using the QiAmp Ultra-Sensitive Virus Kit (Qiagen, Hilden, Germany) and whole genome amplification 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 Illumina HiSeq 2500 platform (Illumina, Inc, San Diego, CA, USA). During that study (5), we identified one participant with a metagenome which consistently showed the presence of a feline papillomavirus and who also self-identified as being an owner of a domesticated house cat.

To further investigate this observation, 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% bp similarity and a k-value of 13 (6). Mapped reads were then processed using KHMER v2.0.0 (7) to remove singletons and overly-abundant reads using k-mer profiles. All filtered papillomavirus reads were then de-novo assembled using MEGAHIT v1.2.8 (8) and the subsequent assembly quality was assessed using QUAST v5.0.2 (9). Read coverage for the assembly was an average of 295.18 with a standard deviation of 46.61 reads per base. To quality check the metagenome assembly, the putative viral contigs were evaluated using CheckV v0.7.0 (10), and classified using 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 cut-off) (12), as well as the protein-coding

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

classification tool Kaiju v1.7 (13). All tools were run with default parameters unless otherwise specified.

A complete circular feline papillomavirus genome assembly was identified which showed 74% sequence similarity to feline papillomavirus type 2 (family *Papillomaviridae*, genus *dyothetapapillomavirus*). Genes of this single virus were annotated with PROKKA v1.14.5 (14) using the embedded viral annotation database and Cenote-Taker 2 v. 2.1 (15). This genome, which we identified as strain P20, was 8069 bp in length and exhibited a GC content of 54.38%. Six open reading frames were annotated, 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) (Figure 1; visualized with SnapGene, GSL Biotech LLC, San Diego, CA; https://snapgene.com). Using BLASTn, our genome assembly showing the greatest similarity (with an average of 70.3% shared amino acids) to a papillomavirus isolated from the skin of a domestic Maine Coon house cat in 2007 (GenBank accession NC038520).

To determine the phylogenetic placement of strain P20 identified here, we downloaded all feline papilloma genomes available on NCBI (along with two outgroups from closely related hosts – domestic dog and Weddell Seal) then 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 (Figure 1) was generated using IQTREE v1.6.12 (18) under the LG+F+I+G4 substitution model optimally

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determined by MODEL-FINDER 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 accessions.

DATA AVAILABILITY

All raw sequencing data has been deposited in the NCBI Short Read Archive (SRA) under the accession code PRJNA754140. Within that project, the 15 sequencing files used in this study were the following: SRS9770471, SRS9770472, SRS9770476, SRS9770495, SRS9770500, SRS9770501, SRS9770587, SRS9770588, SRS9770589, SRS9770614, SRS9770616, SRS9770617, SRS9770642, SRS9770643, SRS9770644. The complete feline papillomavirus P20 genome has been submitted to NCBI Genbank with the accession number of 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.

FUNDING SOURCES

This work was supported by the Department of Justice (grant numbers 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.

DECLARATION OF COMPETING INTERESTS

None to declare.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGMENTS

This work was completed using the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Rector, A. and Van Ranst, M., 2013. Animal papillomaviruses. *Virology*, *445*(1-2), pp.213-223.

2. Rector, A., Lemey, P., Tachezy, R., Mostmans, S., Ghim, S.J., Van Doorslaer, K., Roelke, M., Bush, M., Montali, R.J., Joslin, J. and Burk, R.D., 2007. Ancient papillomavirus-host co-speciation in Felidae. *Genome biology*, *8*(4), pp.1-12.

3. Munday, J.S. and Thomson, N.A., 2021. Papillomaviruses in Domestic Cats. *Viruses*, *13*(8), p.1664.

4. Kaynarcalidan, O. and Oğuzoğlu, T.Ç., 2021. The oncogenic pathways of papillomaviruses. *Veterinary and Comparative Oncology*, *19*(1), pp.7-16.

5. Graham, E.H., Fernando, S., Clarke, J., Herr, J.R. and Adamowicz, M., 2022. The Application of The Skin Virome for Human Identification. *Forensic Science International - Genetics*. doi: https://doi.org/10.1016/j.fsigen.2022.102662

6. Bushnell, B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. https://sourceforge.net/projects/bbmap/

7. Crusoe M.R., Alameldin H.F., Awad S., Boucher E., Caldwell A., Cartwright R., Charbonneau A., Constantinides B., Edvenson 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 J.R., Howe A., Hyer A., Härpfer A., Irber L., Kidd R., Lin D., Lippi J., Mansour T., McA'Nulty P., McDonald E., Mizzi J., Murray K.D., Nahum J.R., Nanlohy K., Nederbragt A.J., Ortiz-Zuazaga H., Ory J., Pell J., Pepe-Ranney C., Russ Z.N., Schwarz E., Scott C., Seaman J., Sievert S., Simpson J., Skennerton C.T., Spencer J., Srinivasan R., Standage D., Stapleton J.A., Steinman S.R., Stein J., Taylor B., Trimble W., Wiencko H.L., Wright M., Wyss B., Zhang Q., zyme e., Brown C.T., 2015. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000 Res*earch. 2015 Sep 25; 4:900. doi: 10.12688/f1000research.6924.1.

8. Li, D., Liu, C.M., Luo, R., Sadakane, K. and Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), pp.1674-1676.

9. Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics 29*(8), pp.1072-1075.

10. Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S. and Kyrpides, N.C., 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature biotechnology*, *39*(5), pp.578-585.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11. Wood, D.E. and Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, *15*(3), pp.1-12.

12. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L., 2009. BLAST+: architecture and applications. *BMC bioinformatics*, *10*(1), pp.1-9.

13. Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature communications*, *7*(1), pp.1-9.

14. Seemann T. 2014. *Prokka: rapid prokaryotic genome annotation*. Bioinformatics 2014 Jul 15;30(14):2068-9.

15. Tisza M.J., Belford A.K., Domínguez-Huerta G., Bolduc B., Buck C.B., 2021. Cenote-Taker 2 democratizes virus discovery and sequence annotation. *Virus Evolution*, Volume 7, Issue 1, January 2021, veaa100, https://doi.org/10.1093/ve/veaa100

16. Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), pp.1792-1797.

17. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540-552.

18. Minh B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., 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 B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 14:587-589. https://doi.org/10.1038/nmeth.4285

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.01.466825; this version posted February 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.
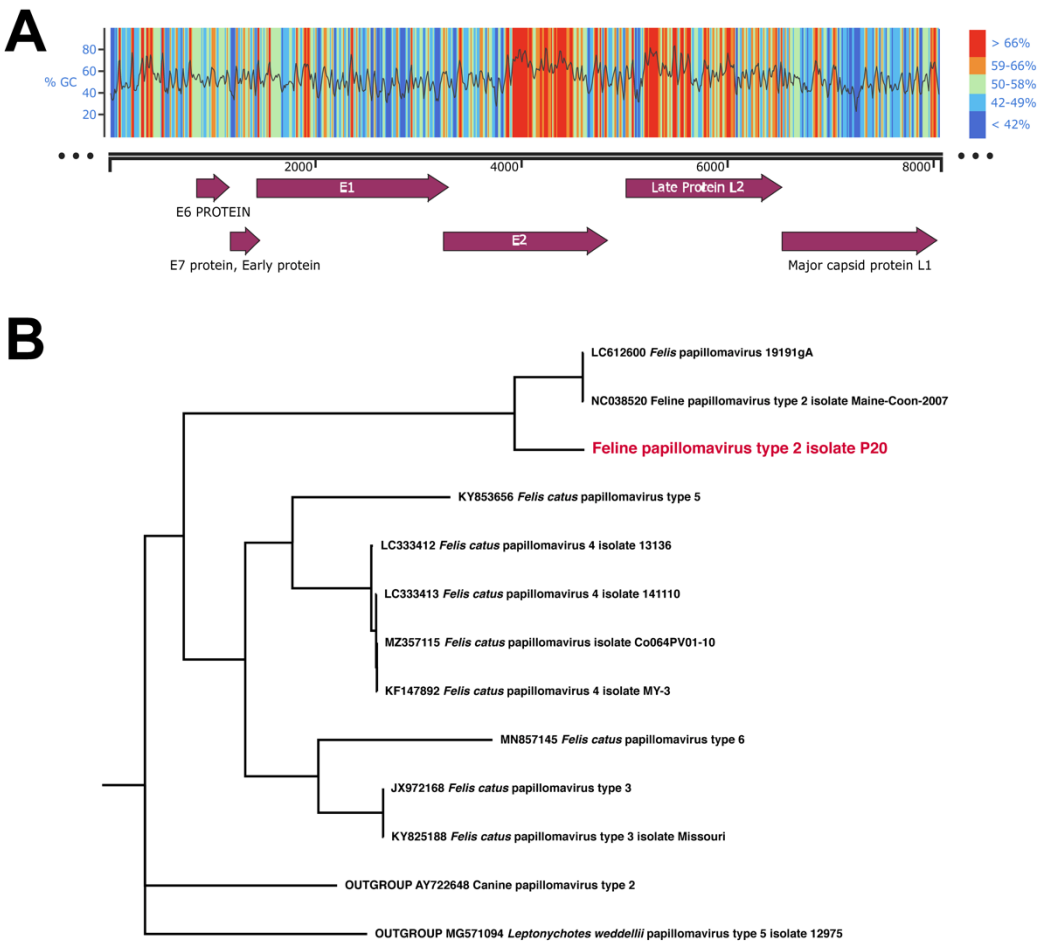
Figure 1. **A.** Linear genome map, 8069 bp in length, of Feline papillomavirus type 2 like virus. Top portion of the figure shows %GC content for regions of the circular genome. Higher GC content are shown in warmer colors (red) and lower GC content are shown in cooler colors (blue). The location of the six gene regions in the genome are shown below which consist of open reading frame annotations (purple arrows) of 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. The tree was rooted with Canine papillomavirus type 2 and *Leptonychotes weddelii* (Wedell Seal) papilloma virus type 5.