Novel Feline Papillomavirus Isolate P20 Assembled from Metagenomic Data Isolated from 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

ABSTRACT

A novel feline papillomavirus isolate was assembled from metagenomic sequencing data collected from the human skin of a house cat owner. This circular papillomavirus isolate P20 is 8069 bp in length and displays genome organization typical of feline papillomaviruses, but only exhibits approximately 75% synteny to other feline papillomaviruses.

ANNOUNCEMENT

Papillomaviruses are most well-known 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, which is conserved across the papilloma viruses (1). The broad phylogenetic relationships of mammalian papilloma viruses suggest that their evolution acutely mirrors their host relationships (2). In recent years, numerous studies have reported on the discovery and diversity of papillomaviruses associated with domestic and wild cats (3,4,5). Currently, seven types of feline papillomaviruses are recognized that may or may not produce both skin and oral lesions (5,6). Multiple studies in the last year have tied feline papilloma viruses to skin and oral squamous cell carcinomas in house cats (7,8,9,10).

In a previous study addressing the diversity and stability of viruses on human skin, we sampled three locations – left hand, right hand, and scalp – over a longitudinal 6-month period (11). Viral enriched metagenomic DNA was extracted from skin swabs and sequenced on an Illumina Hi-Seq sequencer as previously described (11). In that study,

a questionnaire was given to participants to collect pertinent information about subject lifestyle and environmental variables, such that we could understand how the human skin virome is shaped and maintained. Questions on the survey included the frequency and locations of travel, skin care habits, and recent contact with animals (i.e. types of pets). When evaluating viral diversity of the virome samples it was noted that one study participant had a metagenomic virome profile that consistently showed synteny to a feline papilloma virus. That individual had also identified in their questionnaire as being an owner of a domesticated house cat.

To further investigate this observation, a dataset containing the viral metagenomic assembly of compiled data from that specific human individual was generated. This included 15 different metagenomic samples from across three body locations and five sampling times. In our previous study, a combination of laboratory techniques and bioinformatic tools to isolate the viral component of human skin was employed (11). Here, utilizing the virome sample reads from that study, sequencing reads were first mapped to Papillomaviridae reference database material from NCBI using BBMap v38.94 (12) with flags and parameters described for high precision mapping as suggested in the BBMap Guide. Mapped reads were the further processed using KHMER v2.0.0 (13) to remove singleton reads and high copy number reads using k-mer profiles. All filtered papilloma virus reads were then assembled using MEGAHIT v1.2.8 (14) and the subsequent assembly quality was assessed using QUAST v5.0.2 (15). To double check the assembly, the putative viral contigs – designated as contigs containing annotated papilloma virus genes – were identified using the tool CheckV v0.7.0 (16).

80 Contigs were further classified using both nucleotide-based classification tools, such as
81 Kraken2 v2.0.8-beta (17), Demovir (https://github.com/feargalr/Demovir), Blastn (with a
82 >10% query coverage cut-off) (18), and using a classification tool based on protein
83 coding sequences, Kaiju v1.7 (19). Based on these assembly and contig taxonomic
84 identification methods, a feline papilloma virus genome assembly was identified which
85 showed 74% percent similarity to feline papilloma virus type 2. Genes of this single virus
86 were annotated with PROKKA v1.14.5 (20) using the embedded viral annotation
87 database and Cenote-Taker 2 v. 2.1 (21).
88
89 We identified a complete, circular feline papilloma virus that was 8069 bp in length
90 (Figure 1) which we identified as isolate P20. Through annotation using PROKKA
91 v1.14.5, six open reading frames which included the E6 Protein, E7 protein, E1 protein,
92 E2 protein, Late Protein L2, and Major capsid protein L1 were predicted (Figure 2).
93 These six open reading frames are shared among papillomaviruses. Genome maps
94 (Figure 1, Figure 2) were visualized using the SnapGene software (GSL Biotech LLC,
95 San Diego, CA; https://snapgene.com). Using a BLASTn search in NCBI GenBank, we
96 found that the papilloma virus showing the greatest synteny to our genome assembly
97 was isolated from the skin of a domestic Maine Coon house cat in 2007 (GenBank
98 accession number NC038520). Our isolate showed 70.3% amino acid synteny for all
99 coding regions when compared to that accession.
100
101 To determine the phylogenetic placement of isolate P20 identified in this study, we
102 downloaded all papilloma viruses identified as having a feline host on NCBI (along with

103 two outgroups from closely related hosts – domestic dog and Weddell Seal) and
104 annotated open reading frames using PROKKA v1.14.5 and aligned the concatenated
105 amino acid sequences using MUSCLE v3.8.1551 (22) for the six genes (E6 Protein, E7
106 protein, E1 protein, E2 protein, Late Protein L2, and Major capsid protein L1) shared
107 among all the feline papilloma available on NCBI. After removal of uninformative amino
108 acid regions using GBLOCKS v0.91b (23), a phylogenetic tree (Figure 3) of the isolates
109 was generated using IQTREE v1.6.12 (24) under the LG+F+I+G4 substitution model
110 optimally determined by MODEL-FINDER v1.5.4 (25). This phylogenetic tree (Figure 3),
111 visualized with FIGTREE v1.4.4 (https://github.com/rambaut/figtree), placed our isolate,
112 in close proximity to other feline papilloma type 2 accessions.
113

114 **DATA AVAILABILITY**

115 All raw sequencing data has been deposited in the NCBI Short Read Archive (SRA)
116 under the accession code PRJNA754140. The complete feline papillomavirus isolate
117 P20 genome has been submitted to NCBI Genbank with the accession number of
118 OL310516. All metadata, sequences, annotation files, and scripts used here are publicly
119 available and archived at:
120 https://github.com/HerrLab/Graham_2021_feline_papilloma_P20.
121
122

123 **FUNDING SOURCES**

124 This work was supported by the Department of Justice (grant numbers 2017-IJ-CX-
125 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.

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.

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. Sundberg, J.P., Van Ranst, M., Montali, R., Homer, B.L., Miller, W.H., Rowland, P.H., Scott, D.W., England, J.J., Dunstan, R.W., Mikaelian, I. and Jenson, A.B., 2000. Feline papillomas and papillomaviruses. *Veterinary Pathology*, *37*(1), pp.1-10.

4. Munday, J.S., 2014. Papillomaviruses in felids. *The Veterinary Journal*, *199*(3), pp.340-347.

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

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

7. Altamura, G., Cardeti, G., Cersini, A., Eleni, C., Cocumelli, C., Bartolomé del Pino, L.E., Razzuoli, E., Martano, M., Maiolino, P. and Borzacchiello, G., 2020. Detection of *Felis catus* papillomavirus type-2 DNA and viral gene expression suggest active infection in feline oral squamous cell carcinoma. *Veterinary and comparative oncology*, *18*(4), pp.494-501.

8. Carrai M, Van Brussel K, Shi M, Li C-X, Chang W-S, Munday JS, Voss K, McLuckie A, Taylor D, Laws A, Holmes EC, Barrs VR, Beatty JA. 2020. Identification of a Novel Papillomavirus Associated with Squamous Cell Carcinoma in a Domestic Cat. Viruses 12(1):124. https://doi.org/10.3390/v12010124

9. Chu, S., Wylie, T.N., Wylie, K.M., Johnson, G.C., Skidmore, Z.L., Fleer, M., Griffith, O.L. and Bryan, J.N., 2020. A virome sequencing approach to feline oral squamous cell carcinoma to evaluate viral causative factors. *Veterinary microbiology*, *240*, p.108491.

10. Ito, S., Chambers, J.K., Sumi, A., Yamashita-Kawanishi, N., Omachi, T., Haga, T., Nakayama, H. and Uchida, K., 2021. Involvement of *Felis catus* papillomavirus type 2 in the tumorigenesis of feline Merkel cell carcinoma. *Veterinary pathology*, p.03009858211045440.

11. Graham, E.H., Fernando, S., Clarke, J., Herr, J.R. and Adamowicz, M., 2021. The Application of The Skin Virome for Human Identification. *bioRxiv*. doi: https://doi.org/10.1101/2021.09.10.459834

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

13. 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. F1000Res. 2015 Sep 25; 4:900. doi: 10.12688/f1000research.6924.1.

14. 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.

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

16. 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.

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

18. 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.

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

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

21. 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

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

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

24. 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

25. 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

Figure 1. Circular genome map, which is 8069 bp in length, of Feline papillomavirus type 2 like virus. From outside to center: open reading frame annotations of E6 Protein, E7 protein, E1 protein, E2 protein, Late Protein L2, and Major capsid protein L1 (inside purple arrows)

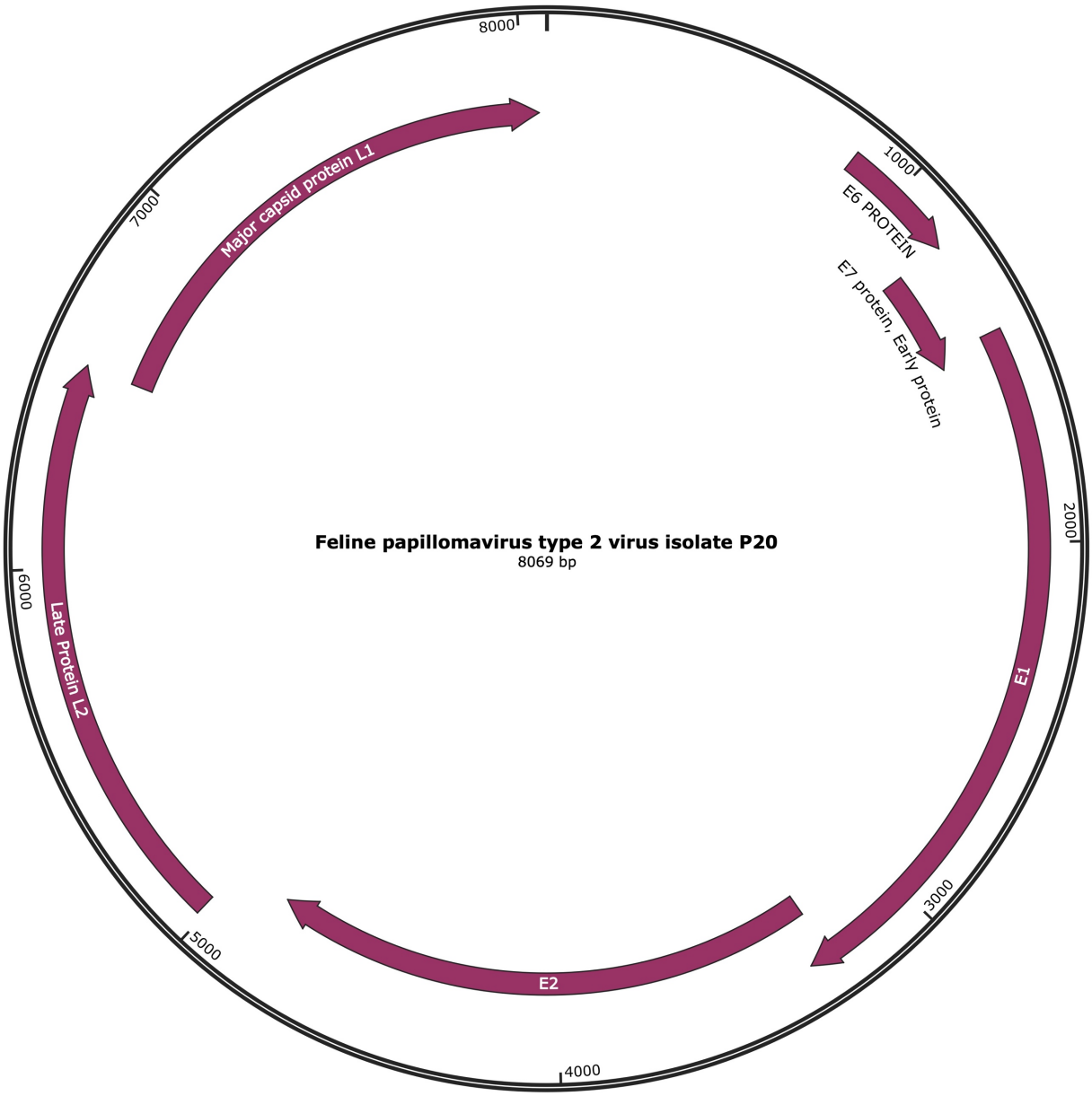


Figure 2. Linear genome map, 8069 bp in length, of Feline papillomavirus type 2 like virus. Top portion of the figure shows %GC content for regions of the circular genome. Higher GC content are shown in warmer colors (red) and lower GC content are shown in cooler colors (blue). Below %GC content is the location of the six gene regions in the viral genome consisting of open reading frame annotations of E6 Protein, E7 protein, E1 protein, E2 protein, Late Protein L2, and Major capsid protein L1 (inside purple arrows)

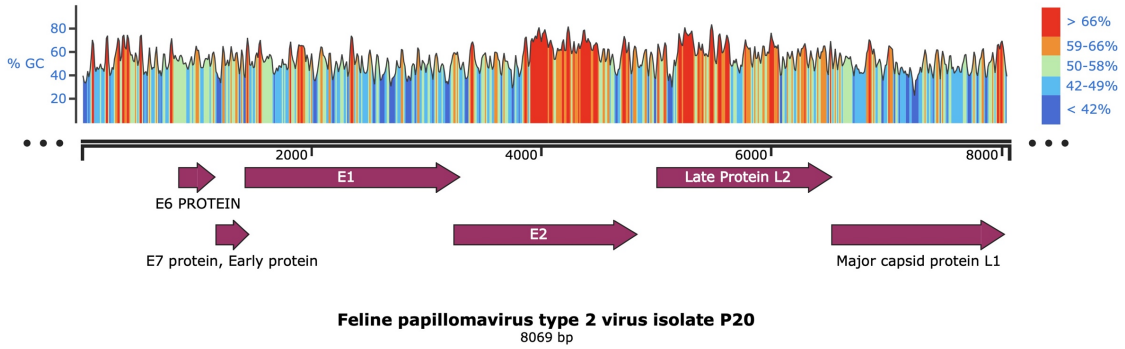
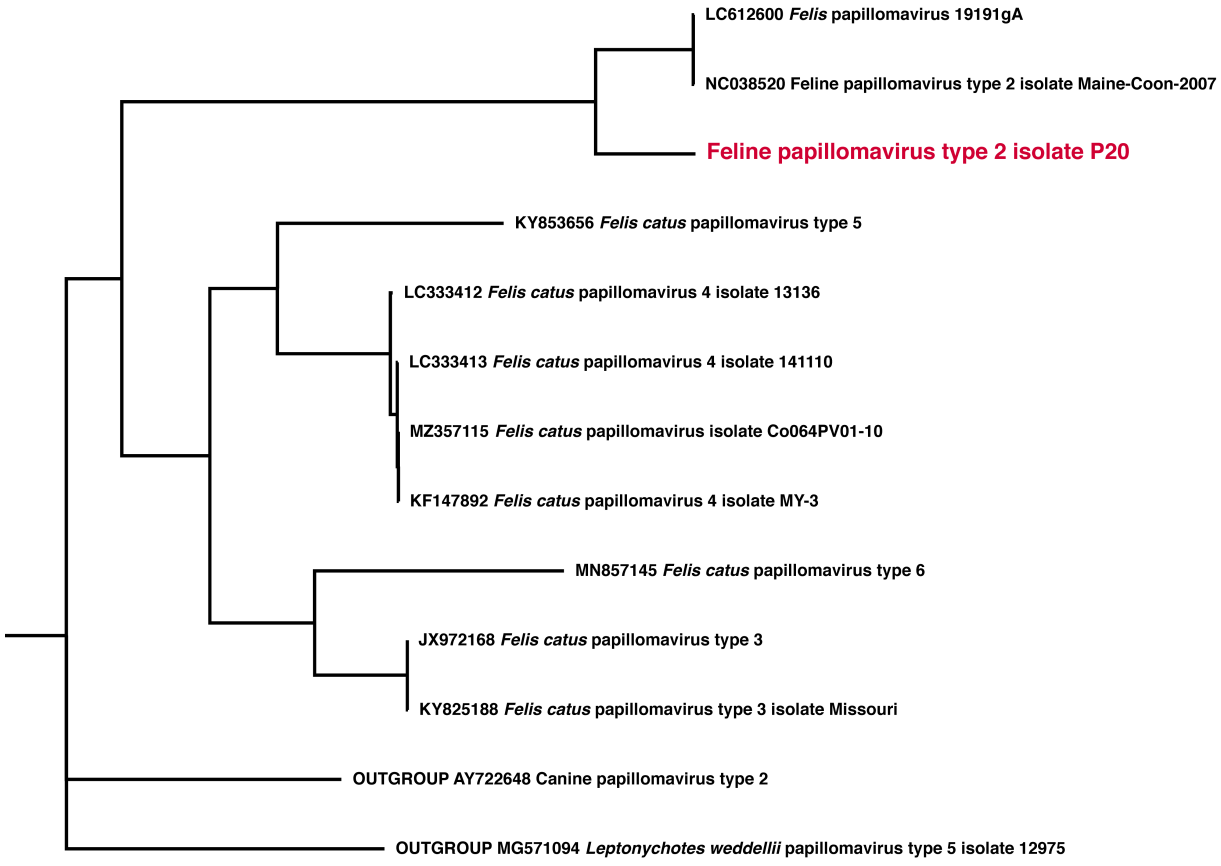
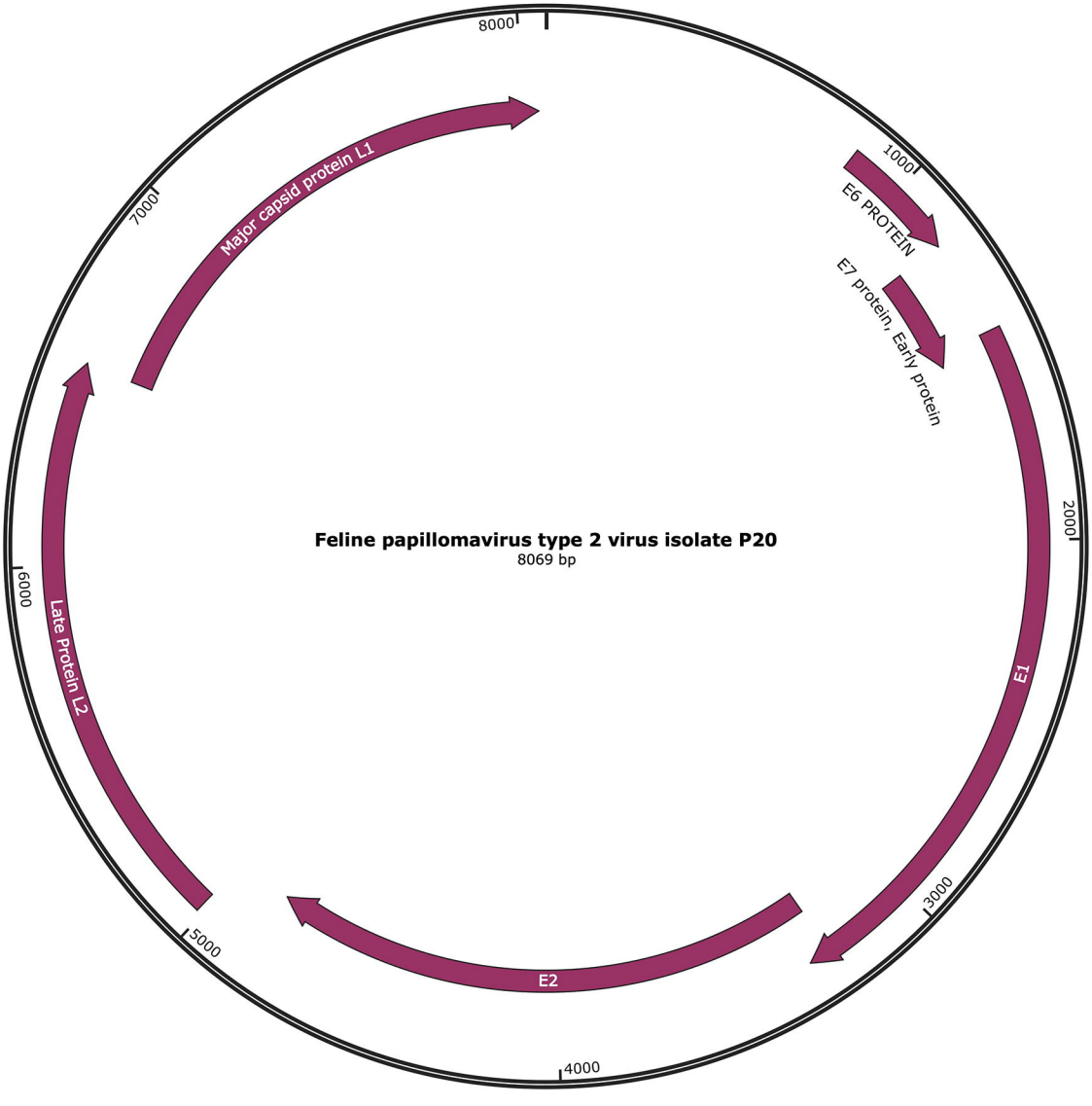
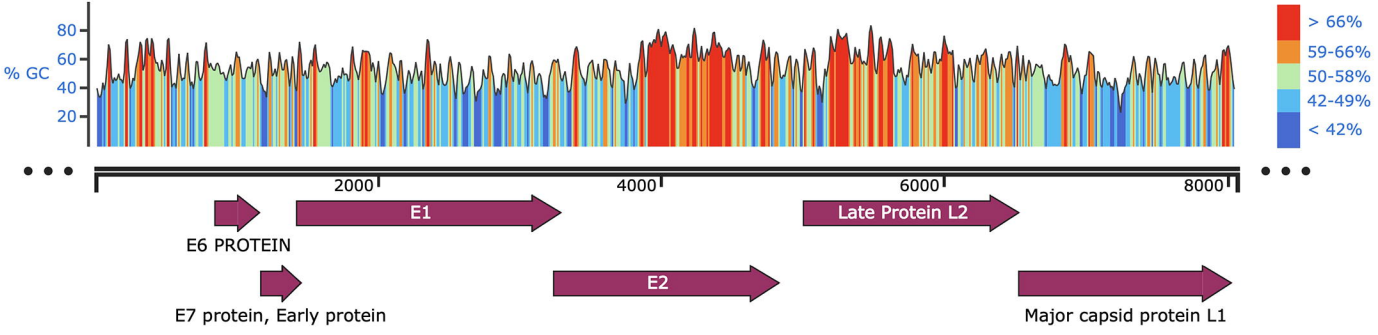


Figure 3. Phylogenetic tree of the placement of the feline type 2 papilloma virus identified in this study along with nine other feline papilloma viruses. The tree was rooted with Canine papilloma virus type 2 and *Leptonychotes weddelii* (Wedell Seal) papilloma virus type 5.

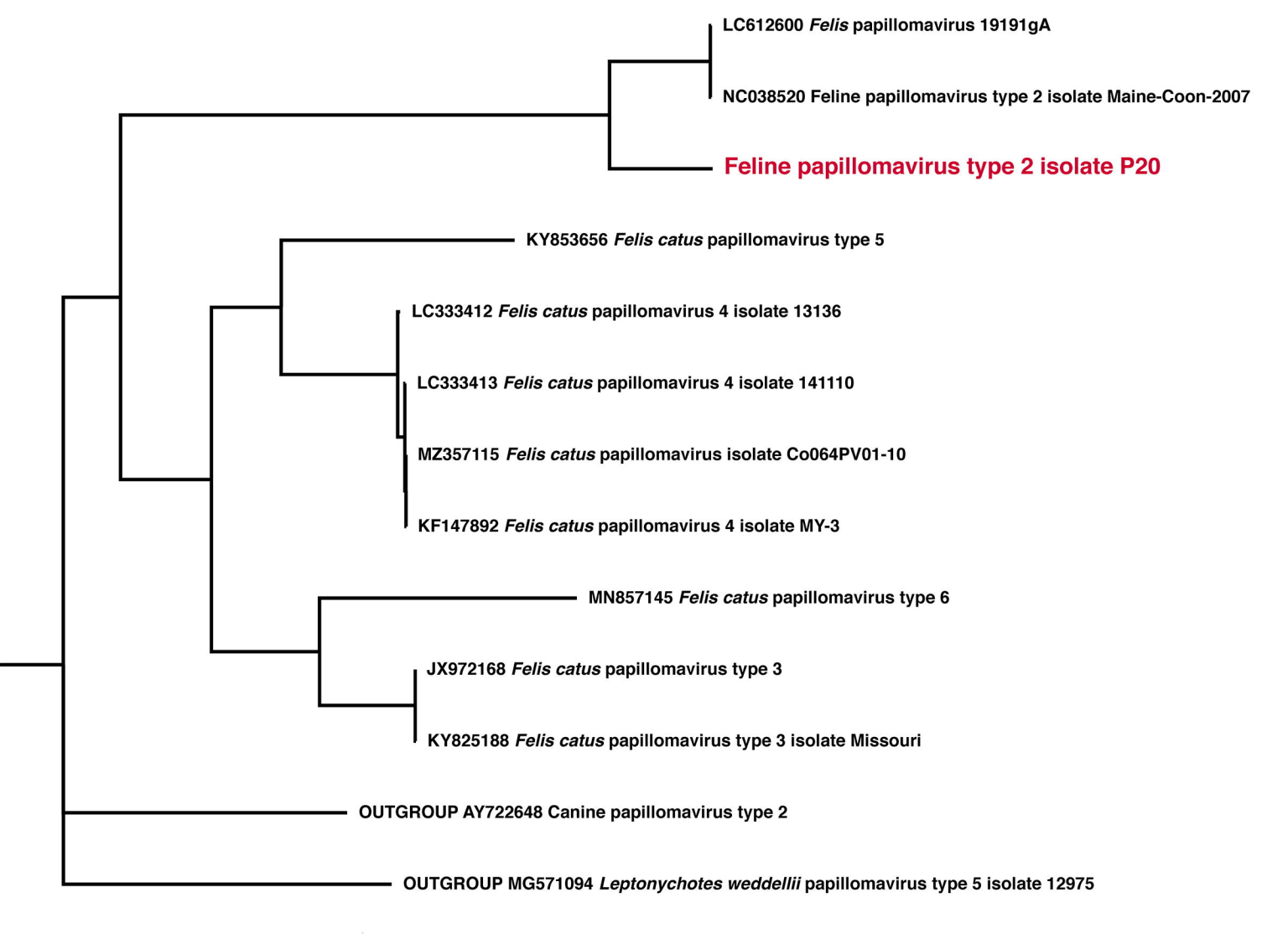



Feline papillomavirus type 2 virus isolate P20
8069 bp
8000
1000
2000
3000
4000
5000
6000
7000
E6 PROTEIN
E7 protein, Early protein
E1
E2
Late Protein L2
Major capsid protein L1



Feline papillomavirus type 2 virus isolate P20

8069 bp


LC612600 Felis papillomavirus 19191gA
NC038520 Feline papillomavirus type 2 isolate Maine-Coon-2007
Feline papillomavirus type 2 isolate P20
KY853656 Felis catus papillomavirus type 5
LC333412 Felis catus papillomavirus 4 isolate 13136
LC333413 Felis catus papillomavirus 4 isolate 141110
MZ357115 Felis catus papillomavirus isolate Co064PV01-10
KF147892 Felis catus papillomavirus 4 isolate MY-3
MN857145 Felis catus papillomavirus type 6
JX972168 Felis catus papillomavirus type 3
KY825188 Felis catus papillomavirus type 3 isolate Missouri
OUTGROUP AY722648 Canine papillomavirus type 2
OUTGROUP MG571094 Leptonychotes weddellii papillomavirus type 5 isolate 12975