




The Polyomavirus Episteme: A Database for Researchers

María José Andrade,^a Gau Shoua Vue,^b  Michael J. Imperiale^b

^aSwarthmore College, Swarthmore, Pennsylvania, USA

^bDepartment of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT *Polyomaviridae* members are DNA viruses that infect a variety of species. Since the first polyomavirus was isolated in 1953, technological advancements have led to the discovery of many polyomaviruses in multiple species. The Polyomavirus Episteme curates data about each polyomavirus, drawing from public databases to present known taxonomic, genomic, and clinical information about polyomaviruses.

The first polyomavirus was isolated in 1953 from tumors in laboratory mice, and the name of the virus was thus composed from the Greek roots poly-, meaning many, and -oma, meaning tumors (1, 2). Polyomaviruses are host specific and have been found in mammals, birds, and some fish, causing symptomatic infections or cancer in some hosts (3). Importantly for humans, they have been known to cause illness in immunocompromised patients. Viruses belonging to the family *Polyomaviridae* have nonenveloped, double-stranded circular DNA genomes that are associated with histones taken from their host (4). The viruses can express five to seven proteins, including three viral capsid proteins (VP1, VP2, and VP3), as well as two proteins involved in cell cycle regulation and replication of DNA, called the large tumor antigen (LT) and small tumor antigen (ST) proteins (3). The transcription of these proteins is temporally regulated; early proteins are involved in replication and are expressed prior to the onset of DNA replication, and late proteins are involved in capsid formation and are expressed after DNA replication has begun.

Polyomaviruses are very diverse. A complex mixture of evolutionary processes related to polyomavirus host specificity and recombination might have led to this diversity (5). In recent years, many novel polyomaviruses have been discovered due to advancements in cloning and sequencing methods (6). These advancements have created a need for an integrated resource with information on all polyomaviruses. An episteme, a Greek term referring to a system of scientific knowledge, captures the purpose of such a project. Following the model of a similar resource created for viruses belonging to the *Papillomaviridae* family (7), the Polyomavirus Episteme (<https://sites.google.com/umich.edu/polyomavirusepisteme>) has been created as a resource for polyomavirus sequence data and analysis, with curated information on all discovered polyomaviruses.

The Polyomavirus Episteme includes taxonomic, genomic, and clinical information about each polyomavirus (Fig. 1A). The data were collected from public sources, including PubMed and the NCBI nucleic acid and protein databases.

Taxonomic data include information about the genus, species, and natural host of each virus. The genus of a polyomavirus is determined from phylogenetic relationships of LT proteins (8). There are four genera, each with a restricted natural host range, and there are more than 100 species. Some viruses have not yet been assigned to a genus and are classified under “unknown genus.”

Genomic data include data on the early and late proteins expressed by a specific reference strain of each virus. Polyomavirus genomes typically encode two regulatory proteins during early infection, the LT and ST proteins, which are called the early proteins. The three capsid proteins, VP1, VP2, and VP3, are expressed after the onset of

Citation Andrade MJ, Vue GS, Imperiale MJ. 2021. The Polyomavirus Episteme: a database for researchers. *Microbiol Resour Announc* 10:e00108-21. <https://doi.org/10.1128/MRA.00108-21>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2021 Andrade 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 Michael J. Imperiale, imperial@umich.edu.

Received 1 February 2021

Accepted 15 March 2021

Published 8 April 2021

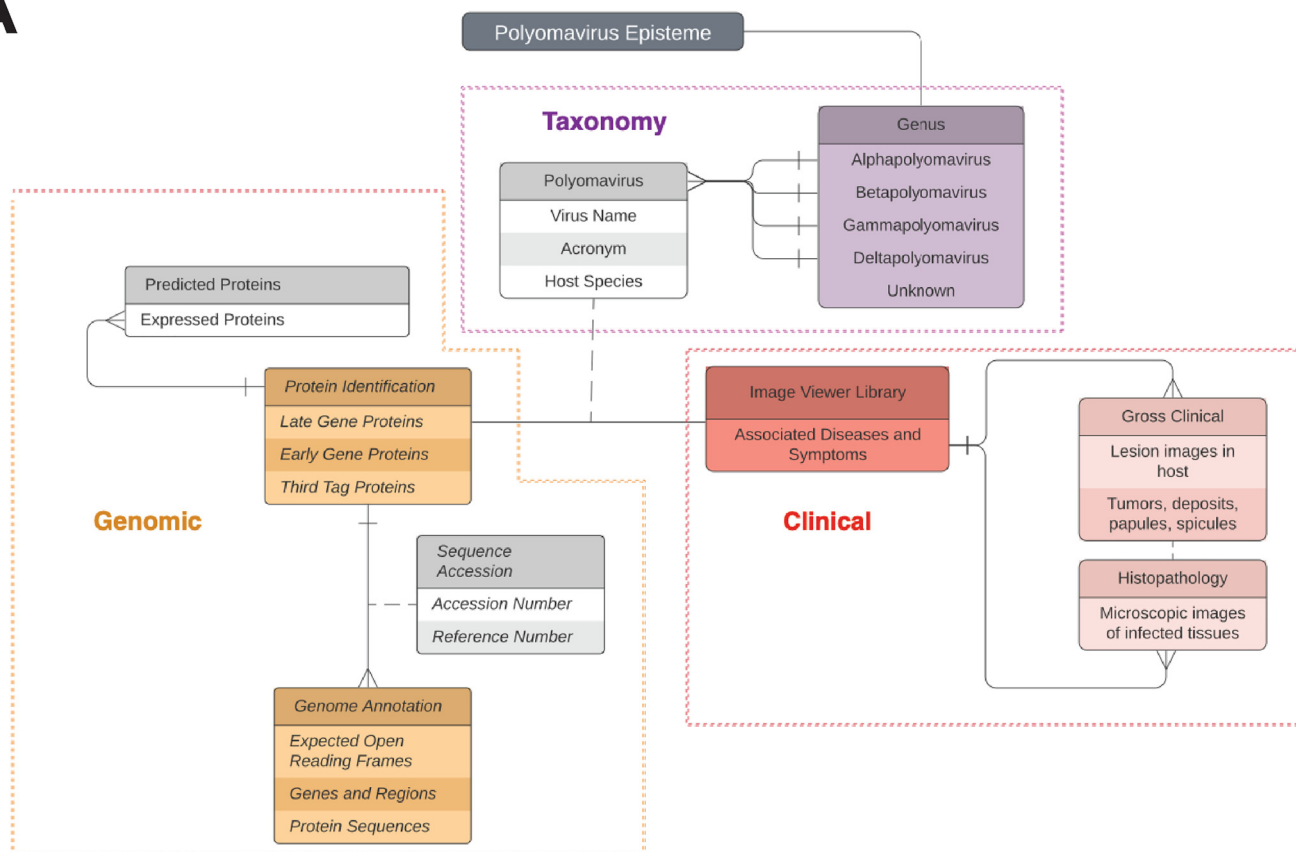
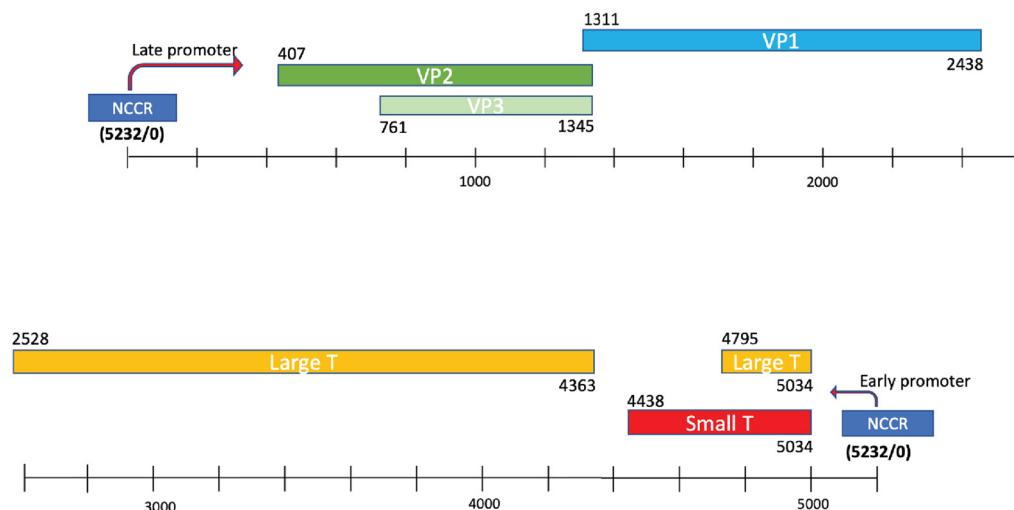
A**B**

FIG 1 (A) Structure of the database, presented as an entity-relationship diagram. (B) Example of an annotated late coding region, from trichodysplasia spinulosa-associated polyomavirus.

viral DNA replication and are thus called late proteins. VP1 enables attachment and entry of virus into host cells, and VP2 and VP3, which have overlapping open reading frames (ORFs) (with VP3 being identical to the C terminus of VP2), have nuclear localization signals that, once virion disassembly is initiated in the cytosol, guide the particles into the nucleus, where further disassembly occurs (3, 9). Most polyomaviruses have these ORFs, but many polyomaviruses express additional early and late proteins. The genomes of most polyomaviruses are approximately 5,000 bp and have a

noncoding control region (NCCR) located between the early and late regions, which contains the origin of DNA replication and the early and late promoters (9).

Through analysis of the ORFs of each virus, which have been curated from public databases such as the NCBI databases and from relevant publications, each viral genome has been annotated to show the late and early regions. Encoded proteins and some important motifs are annotated with their genome locations by base pair. The gaps in the annotation of LT proteins and some other proteins in certain polyomaviruses indicate spliced mRNA. The NCCR has high sequence variability among polyomaviruses, and it is not included in the annotation of all viruses. An example of annotation is shown in Fig. 1B.

Polyomaviruses can cause symptomatic infections and tumors in human and veterinary hosts. The Polyomavirus Episteme includes a library of gross clinical and histopathology images of lesions and diseases associated with each polyomavirus. The relevant publications are specified under each image.

The Polyomavirus Episteme uses Google Sites, with underlying code written in HTML. The annotations were manually edited on the basis of the NCBI databases and were confirmed using additional publications in the scientific literature to ensure that all expected ORFs were included.

The Polyomavirus Episteme has been curated to aid the study of polyomaviruses by integrating and linking multiple resources and facilitating navigation through the growing number of discoveries of novel virus types. As interests in polyomavirus biology and nonhuman infections grow, the Polyomavirus Episteme can become a resource for facilitating the annotation and classification of novel viral genomes. We hope to periodically update and expand the database annotation of genomes to include transcriptional promoters, enhancers, and binding sites for key *trans*-acting factors. We also hope to receive feedback from people in the field who may have recommendations for useful features that could be incorporated in the future.

Data availability. The episteme can be found at <https://sites.google.com/umich.edu/polyomavirusepisteme>.

ACKNOWLEDGMENTS

We acknowledge Ryan Echlin and Deb McCaffrey of the Health Information Technology and Services (HITS) academic information technology department at the University of Michigan for their support in creating the website.

We acknowledge the National Summer Undergraduate Research Project (NSURP) for supporting M.J.A. and facilitating this research internship. This work was funded by NIH grant R01 AI060584 awarded to M.J.I.

REFERENCES

- Gross L. 1953. A filterable agent, recovered from Ak leukemic extracts, causing salivary gland carcinomas in C3H mice. *Proc Soc Exp Biol Med* 83:414–421. <https://doi.org/10.3181/00379727-83-20376>.
- Stewart SE, Eddy BE, Gochenour AM, Borgese NG, Grubbs GE. 1957. The induction of neoplasms with a substance released from mouse tumors by tissue culture. *Virology* 3:380–400. [https://doi.org/10.1016/0042-6822\(57\)90100-9](https://doi.org/10.1016/0042-6822(57)90100-9).
- Moens U, Calvignac-Spencer S, Lauber C, Ramqvist T, Feltkamp MCW, Daugherty MD, Verschoor EJ, Ehlers B, ICTV Report Consortium. 2017. ICTV virus taxonomy profile: *Polyomaviridae*. *J Gen Virol* 98:1159–1160. <https://doi.org/10.1099/jgv.0.000839>.
- Hurdiss DL, Morgan EL, Thompson RF, Prescott EL, Panou MM, Macdonald A, Ranson NA. 2016. New structural insights into the genome and minor capsid proteins of BK polyomavirus using cryo-electron microscopy. *Structure* 24:528–536. <https://doi.org/10.1016/j.str.2016.02.008>.
- Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, Katz JP, Pipas JM, McBride AA, Camus AC, McDermott AJ, Dill JA, Delwart E, Ng TF, Farkas K, Austin C, Kraberger S, Davison W, Pastrana DV, Varsani A. 2016. The ancient evolutionary history of polyomaviruses. *PLoS Pathog* 12:e1005574. <https://doi.org/10.1371/journal.ppat.1005574>.
- DeCaprio JA, Garcea RL. 2013. A cornucopia of human polyomaviruses. *Nat Rev Microbiol* 11:264–276. <https://doi.org/10.1038/nrmicro2992>.
- Van Doorslaer K, Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, Sun Q, Kaur R, Huyen Y, McBride AA. 2017. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res* 45:D499–D506. <https://doi.org/10.1093/nar/gkw879>.
- Calvignac-Spencer S, Feltkamp MC, Daugherty MD, Moens U, Ramqvist T, John R, Ehlers B, Polyomaviridae Study Group of the International Committee on Taxonomy of Viruses. 2016. A taxonomy update for the family *Polyomaviridae*. *Arch Virol* 161:1739–1750. <https://doi.org/10.1007/s00705-016-2794-y>.
- DeCaprio JA, Imperiale MJ, Major EO. 2013. Polyomaviruses, p 1633–1661. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), *Fields virology*, 6th ed. Lippincott Williams & Williams, Philadelphia, PA.