



# Pacmanvirus S19, the Second Pacmanvirus Isolated from Sewage Waters in Oran, Algeria

Khalil Geballa-Koukoulas,<sup>a,b</sup> Souhila Abdi,<sup>a</sup>  Bernard La Scola,<sup>a</sup> Guillaume Blanc,<sup>b</sup>  Julien Andreani<sup>a</sup>

<sup>a</sup>MEPHI, APHM, IRD 198, Aix Marseille Université, IHU-Méditerranée Infection, Marseille, France

<sup>b</sup>Aix Marseille Université, Université de Toulon, CNRS, IRD, MIO UM 110, Marseille, France

**ABSTRACT** *Acanthamoeba castellanii* is an amoeba host that was used to isolate a novel strain named pacmanvirus S19. This isolate is the second strain reported and belongs to the extended *Asfarviridae* family. Pacmanvirus S19 harbors a 418,588-bp genome, with a GC content of 33.20%, which encodes 444 predicted proteins and a single Ile-tRNA.

In 2017, the first pacmanvirus (strain A23) was isolated from *Acanthamoeba castellanii* (1) and nested phylogenetically within the extended *Asfarviridae* clade (2). This clade contains the African swine fever virus, a virus with an endemic background that causes swine disease and death (3–7). Faustoviruses (8), kaumoebaviruses (9), abalone asfarvirus-like virus (10), and asfarvirus metagenome-assembled genomes (MAGs) (11) complete the known diversity of this clade. Here, we report the genome sequence of pacmanvirus S19, which was isolated from a sewage sample collected in Cap Falcon, Oran, Algeria (35°46′15.3″N, 0°47′47.2″W), and stored at 4°C before analysis. This virus was isolated using a coculture technique on a 24-well plate, as described by Andreani et al. (1); viral DNA was extracted with an EZ1 Advanced XL automated system (Qiagen, France). A 2 × 251-bp paired-end sequencing strategy was used, and limited-cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on beads, the library was normalized according to the Nextera XT protocol (Illumina) before sequencing on an Illumina MiSeq instrument (8) in a 39-hour single run.

Sequencing yielded 2,586,744 raw reads, which were trimmed and quality controlled by AlienTrimmer (12) (with parameters  $p = 80$ ,  $l = 100$ , and  $k = 10$ ) before *de novo* assembly with SPAdes v 3.11.1 (13) with k-mer sizes of 21, 55, 77, 99, and 127. Remaining gaps were closed by subassembly of reads aligned with HISAT (14) onto orthologous A23 genomic regions corresponding to the gap and its surroundings (500 bp on both ends), as found by BLASTN (15). A linear contig of 418,588 bp (average coverage, 270×), with a GC content of 33.20%, was generated. GeneMarkS (16) predicted 505 genes using the virus option; 61 of those genes were discarded from the final annotation because they were shorter than 300 bp and had no detectable hits in the nonredundant database (BLASTP E values of  $<1E-05$ ; pacmanvirus A23 hits were excluded). The 444 predicted open reading frames (ORFs) were functionally annotated according to the best similarities against two protein databases and two motif databases, in the following order: Swiss-Prot and UniRef90 were searched using BLASTP (E values of  $<1E-05$ ), excluding pacmanvirus A23 hits; Pfam-A and InterPro motif databases were searched using PfamScan (17) and InterProScan (18), respectively (E values of  $<1E-05$ ). Proteins that did not yield detectable hits were annotated as hypothetical (Fig. 1A). Furthermore, one Ile-tRNA (Fig. 1B) was found by both ARAGON (19) and tRNAscan-SE (20). Pacmanvirus tRNAs were also found in some asfarvirus MAGs (11). For comparison, pacmanvirus A23 was reported to have a smaller genome (395,405 bp) containing 465 predicted protein genes and an Ile-tRNA gene, with a GC content of 33.62% (1). The average nucleotide identity (ANI) between the two strains, as calculated by OrthoANIu (21), was 84.97%.

**Citation** Geballa-Koukoulas K, Abdi S, La Scola B, Blanc G, Andreani J. 2021. Pacmanvirus S19, the second pacmanvirus isolated from sewage waters in Oran, Algeria. *Microbiol Resour Announc* 10:e00693-21. <https://doi.org/10.1128/MRA.00693-21>.

**Editor** Kenneth M. Stedman, Portland State University

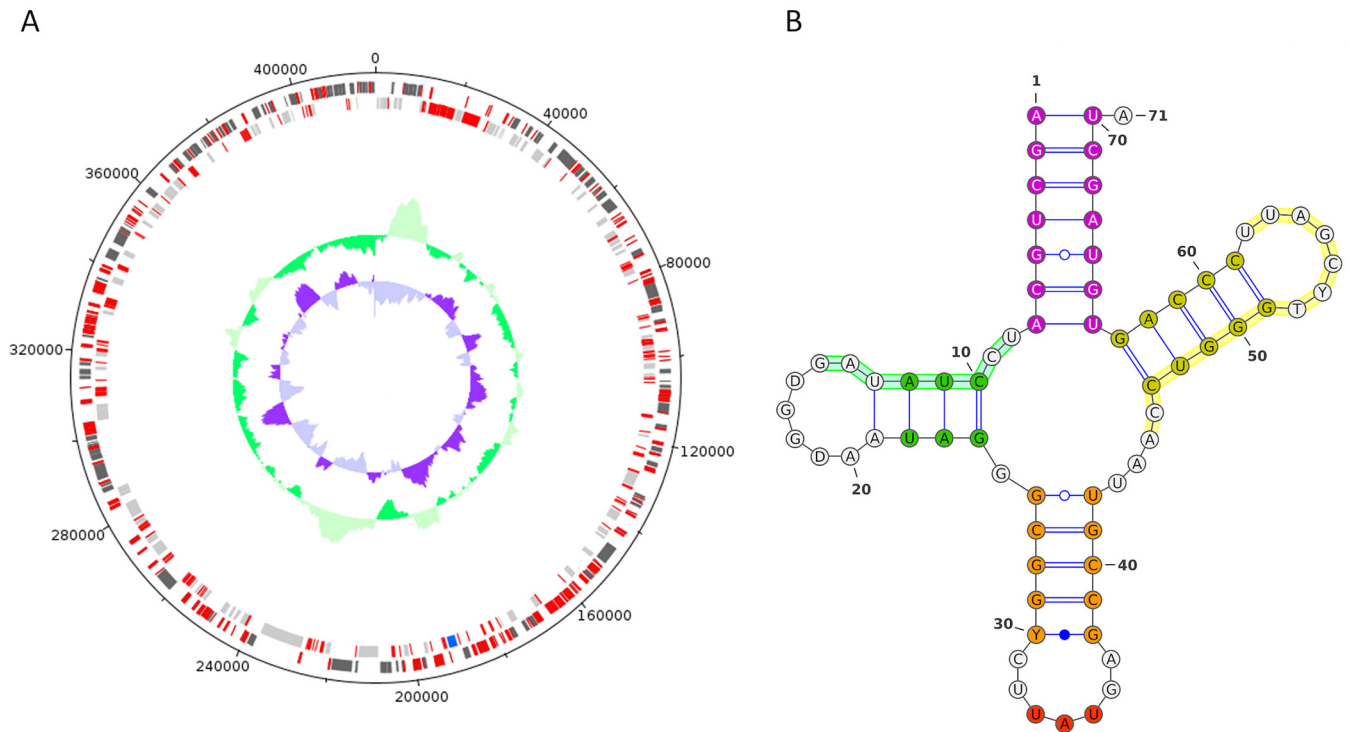
**Copyright** © 2021 Geballa-Koukoulas et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Julien Andreani, [miaguiaibidou@gmail.com](mailto:miaguiaibidou@gmail.com), or Guillaume Blanc, [guillaume.blanc@mio.osupytheas.fr](mailto:guillaume.blanc@mio.osupytheas.fr).

**Received** 2 August 2021

**Accepted** 23 September 2021

**Published** 21 October 2021



**FIG 1** Circular representation of the pacmanvirus S19 genome with its tRNA. (A) The major capsid protein (MCP) gene is marked with a blue rectangle. Hypothetical proteins are highlighted in red. The outer ring corresponds to the positive strand of the genome, while the inner ring corresponds to the negative strand. GC skew is indicated by the small inner purple plot in the center of the genome and the GC content by the outer green plot. Construction of the plot was performed via DNAPlotter (22). (B) The Ile anticodon is marked in red. Orange represents the anticodon stem. Purple corresponds to the amino-acyl stem, while green represents the D loop stem and yellow represents the TΨC stem. D loop and TΨC signals are highlighted with light green and yellow lines, respectively. For the construction of tRNA, the predicted sequence was uploaded to the tRNAMod online tool (23). Information about the tRNA annotation was obtained using ARAGORN (19). The final representation was prepared using VARNA (24).

**Data availability.** Genome and SRA data have been deposited in GenBank under the accession numbers [MZ440852](#) and [SRR15690446](#), respectively.

#### ACKNOWLEDGMENT

This work was supported by the French Government under the Investissements d'Avenir (Investments for the Future) program, managed by the Agence Nationale de la Recherche (National Agency for Research) (reference number Méditerranée 69 Infection 10-IAHU-03).

#### REFERENCES

- Andreani J, Khalil JYB, Sevvana M, Benamar S, Di Pinto F, Bitam I, Colson P, Klose T, Rossmann MG, Raoult D, La Scola B. 2017. Pacmanvirus, a new giant icosahedral virus at the crossroads between *Asfarviridae* and faustoviruses. *J Virol* 91:e00212-17. <https://doi.org/10.1128/JVI.00212-17>.
- Koonin EV, Yutin N. 2019. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv Virus Res* 103:167–202. <https://doi.org/10.1016/bs.aivir.2018.09.002>.
- Niederwerder M. 2021. Risk and mitigation of African swine fever virus in feed. *Animals* 11:792. <https://doi.org/10.3390/ani11030792>.
- Fekede RJ, HaoNing W, Hein VG, XiaoLong W. 2021. Could wild boar be the trans-Siberian transmitter of African swine fever? *Transbound Emerg Dis* 68:1465–1475. <https://doi.org/10.1111/tbed.13814>.
- Bosch-Camós L, López E, Rodríguez F. 2020. African swine fever vaccines: a promising work still in progress. *Porcine Health Manag* 6:17. <https://doi.org/10.1186/s40813-020-00154-2>.
- Hakizimana JN, Yona C, Kamana O, Nauwynck H, Misinzo G. 2021. African swine fever virus circulation between Tanzania and neighboring countries: a systematic review and meta-analysis. *Viruses* 13:306. <https://doi.org/10.3390/v13020306>.
- Bellini S, Casadei G, De Lorenzi G, Tamba M. 2021. A review of risk factors of African swine fever incursion in pig farming within the European Union scenario. *Pathogens* 10:84. <https://doi.org/10.3390/pathogens10010084>.
- Geballa-Koukoulas K, Boudjemaa H, Andreani J, La Scola B, Blanc G. 2020. Comparative genomics unveils regionalized evolution of the faustovirus genomes. *Viruses* 12:577. <https://doi.org/10.3390/v12050577>.
- Geballa-Koukoulas K, Andreani J, La Scola B, Blanc G. 2021. The kaumoeavirus LCC10 genome reveals a unique gene strand bias among “extended *Asfarviridae*.” *Viruses* 13:148. <https://doi.org/10.3390/v13020148>.
- Matsuyama T, Takano T, Nishiki I, Fujiwara A, Kiryu I, Inada M, Sakai T, Terashima S, Matsuura Y, Isowa K, Nakayasu C. 2020. A novel asfarvirus-like virus identified as a potential cause of mass mortality of abalone. *Sci Rep* 10:4620. <https://doi.org/10.1038/s41598-020-61492-3>.
- Karki S, Moniruzzaman M, Aylward FO. 2021. Comparative genomics and environmental distribution of large dsDNA viruses in the family *Asfarviridae*. *Front Microbiol* 12:657471. <https://doi.org/10.3389/fmicb.2021.657471>.
- Criscuolo A, Brisse S. 2013. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* 102:500–506. <https://doi.org/10.1016/j.ygeno.2013.07.011>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
14. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–360. <https://doi.org/10.1038/nmeth.3317>.
  15. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biol Direct* 7:12. <https://doi.org/10.1186/1745-6150-7-12>.
  16. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29: 2607–2618. <https://doi.org/10.1093/nar/29.12.2607>.
  17. Mistry J, Bateman A, Finn RD. 2007. Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* 8:298. <https://doi.org/10.1186/1471-2105-8-298>.
  18. Zdobnov EM, Apweiler R. 2001. InterProScan: an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848. <https://doi.org/10.1093/bioinformatics/17.9.847>.
  19. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
  20. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* 1962:1–14. [https://doi.org/10.1007/978-1-4939-9173-0\\_1](https://doi.org/10.1007/978-1-4939-9173-0_1).
  21. Yoon S-H, Ha S, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>.
  22. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25: 119–120. <https://doi.org/10.1093/bioinformatics/btn578>.
  23. Panwar B, Raghava GP. 2014. Prediction of uridine modifications in tRNA sequences. *BMC Bioinformatics* 15:326. <https://doi.org/10.1186/1471-2105-15-326>.
  24. Darty K, Denise A, Ponty Y. 2009. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25:1974–1975. <https://doi.org/10.1093/bioinformatics/btp250>.