AMERICAN SOCIETY FOR MICROBIOLOGY
Microbiology® Resource Announcements

Closed, Circular Genome Sequence of *Aureococcus anophagefferens* Virus, a Lytic Virus of a Brown Tide-Forming Alga

Alexander R. Truchon,[a] Eric R. Gann,[a*] Steven W. Wilhelm[a]

[a]Department of Microbiology, University of Tennessee, Knoxville, Tennessee, USA

ABSTRACT Here, we report the genomic sequence of *Aureococcus anophagefferens* virus, assembled into one circular contig from both Nanopore and Illumina reads. The genome is 381,717 bp long with a GC content of 29.1%, which includes an additional 5-kb region between the previously predicted polar ends of the reference genome.

We sequenced the genome of *Aureococcus anophagefferens* virus (AaV), a member of the family *Mimiviridae*, within the phylum *Nucleocytoviricota*. AaV has been propagated on its pelagophyte host, *Aureococcus anophagefferens*, since its isolation in the early 2000s and was first sequenced and assembled using Illumina reads in 2014 (1–3). Since then, the virus has been maintained via coculture with *A. anophagefferens* CCMP 1984. The original assembly was predicted to have terminal ends rich in leucine repeat-containing coding sequences (3, 4). As repetitive regions can lead to improper assemblies when exclusively employing short-read sequencing exclusively, we resequenced AaV using short- and long-read sequencing (5).

Viral DNA was extracted from a lysed, xenic *A. anophagefferens* culture as described by Truchon et al. (6). Briefly, virions were concentrated using tangential flow filtration and ultracentrifugation to enrich for virus particles. Particles were digested in agarose CHEF plug molds (Bio-Rad, Hercules, CA, USA) with proteinase K before being run on a low-melting point agarose gel. High-molecular-weight DNA was excised from the gel and purified using a phenol-chloroform method (6).

Long-read sequencing was performed using the Oxford Nanopore Technologies (ONT; Oxford, UK) platform. Genomic DNA libraries generated using the ligation sequencing kit (ONT) were sequenced on a MinION R9.4 flow cell (ONT), producing 284,000 reads that averaged 2,503 bp. Bases were called using the Guppy version 3.0.3 base caller using the config file dna_r9.4.1_450bps_fast.cfg (7). Adapter sequences were removed using Porechop version 0.2.4 (8), and the reads were trimmed with a quality score of 9 and a minimum length of 500 bp using NanoFilt version 2.7.1 (9). The reads were aligned to the AaV reference genome using BBMap version 38.90 (10) and used in the assembly, performed using Canu version 2.1 (11).

DNA was also extracted for short-read sequencing by treating concentrated virions with proteinase K for 1 h at 37°C and extraction via standard phenol-chloroform methods (12). The DNA library was prepared and sequenced on a NextSeq 2000 instrument (Illumina, San Diego, CA, USA) by the Microbial Genome Sequencing Center, generating 12,378,846 reads in 150-bp paired-end format. The Illumina short reads were trimmed for quality using the default settings in CLC Genomics Workbench (Qiagen, Hilden, Germany) and mapped to the Canu-assembled contig using Bowtie 2 version 2.2.3 (13). The assembly was polished with the Illumina reads using Pilon version 1.23 (14), which generated a closed, circular contig of 381,717 bp. The quality and completeness were assessed using CheckV (15). Coding sequences and tRNAs were predicted using Prodigal version 2.6.3 (16) and tRNAscan-SE version 2.0 (17), respectively.

Editor John J. Dennehy, Queens College CUNY

Address correspondence to Steven W. Wilhelm, wilhelm@utk.edu.

*Present address: Eric R. Gann, The Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA.

The authors declare no conflict of interest.

Received 6 April 2022
Accepted 20 May 2022

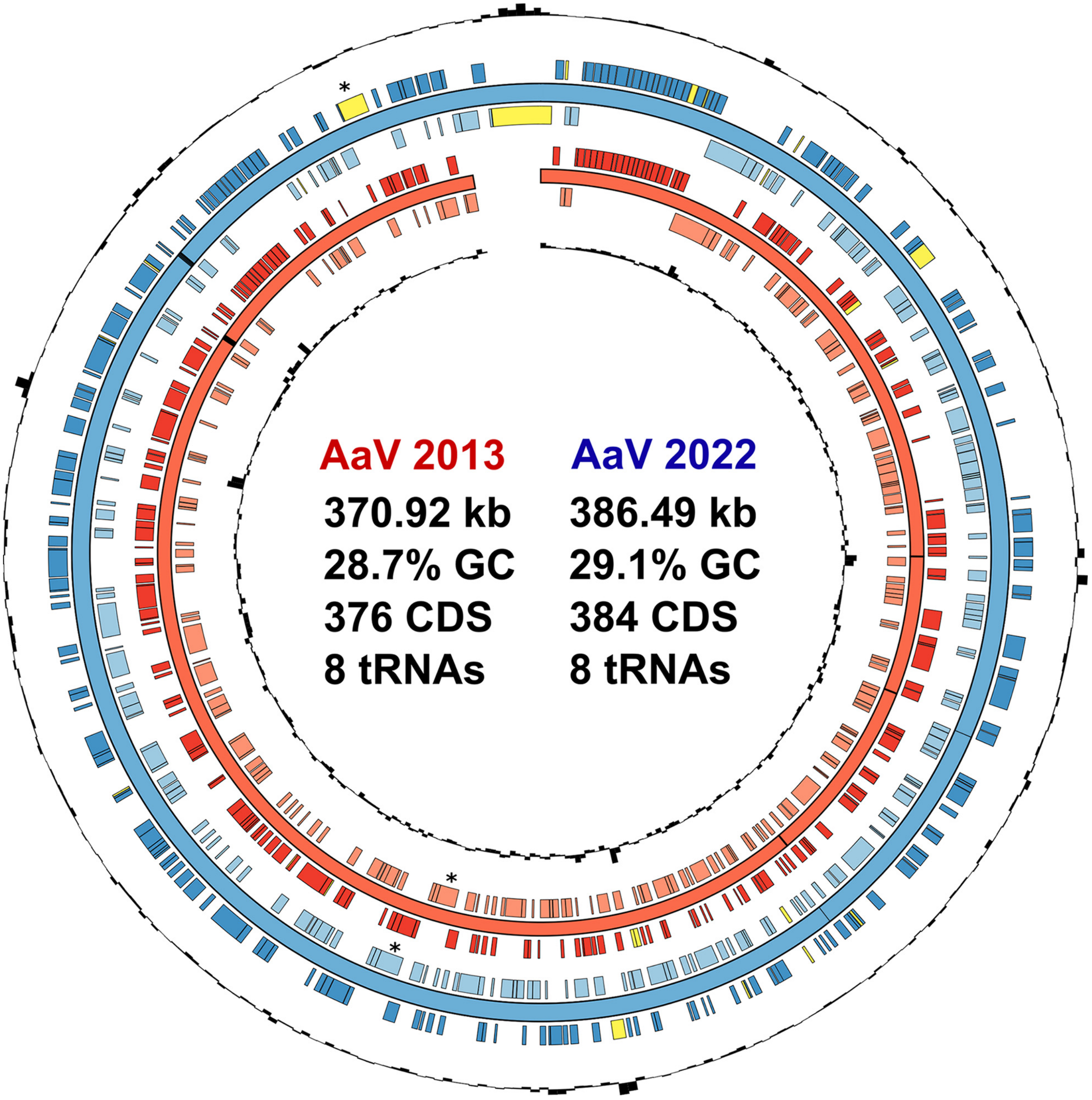

FIG 1 Genomic maps (created in Circos [20]) comparing the original AaV assembly (i.e., 2013) to the updated assembly (i.e., 2022). The complete genome sequences of the two assemblies were aligned, and the coding densities were compared. The rings (from inner to outer) indicate the GC content (2013), minus-strand coding sequences (2013), genomic sequence (2013), plus-strand coding sequences (2013), minus-strand coding sequences (2022), genomic sequence (2022), plus-strand coding sequences, and GC content (2022). Coding sequences present in only one genome are highlighted in yellow, and tRNAs are represented on the genomic sequence (2022) by black bands. Novel and known Rpb2 coding sequences are marked with an asterisk.

A total of 384 coding sequences (CDS) were predicted. Functions were predicted from translated amino acid sequences using the eggNOG-mapper Web server (18). One novel CDS encodes for a putative DNA-dependent RNA polymerase subunit (Rpb2). Only 1 of the 11 RNA polymerase genes encoded by AaV is homologous to the novel subunit, though the two sequences have an amino acid identity below 30%. This supports the hypothesis that two copies of this gene arose from an ancestral duplication among mimiviruses of eukaryotic phototrophs (19). Among other changes to the genome are apparent duplications, gene elongations, and gene combinations missed during the initial assembly (Fig. 1).

Data availability. The raw data and the assembled genome have been indexed at NCBI under the BioProject accession number PRJNA809211. The assembled genome has been assigned the GenBank accession number OM876856.1. The raw MinION and Illumina reads have been archived under the Sequence Read Archive accession numbers SRR16764708 to SRR16764709.

ACKNOWLEDGMENTS

We thank Gary LeCleir, Emily Chase, Robbie Martin, Brittany Zepernick, and Naomi Gilbert for assistance.

This study was supported by funds from the National Science Foundation (IOS-1922958) and the Simons Foundation (735077).

REFERENCES

1. Rowe JM, Dunlap JR, Gobler CJ, Anderson OR, Gastrich MD, Wilhelm SW. 2008. Isolation of a non-phage-like lytic virus infecting *Aureococcus anophagefferens*. J Phycol 44:71–76. https://doi.org/10.1111/j.1529-8817.2007.00453.x.
2. Gobler CJ, Anderson OR, Gastrich MD, Wilhelm SW. 2007. Ecological aspects of viral infection and lysis in the harmful brown tide alga Aureococcus anophagefferens. Aquat Microb Ecol 47:25–36. https://doi.org/10.3354/ame047025.
3. Moniruzzaman M, LeCleir GR, Brown CM, Gobler CJ, Bidle KD, Wilson WH, Wilhelm SW. 2014. Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. Virology 466–467:60–70. https://doi.org/10.1016/j.virol.2014.06.031.
4. Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV, Zhang Y, Collier JL, Wurch LL, Kustka AB, Dill BD, Shah M, VerBerkmoes NC, Kuo A, Terry A, Pangilinan J, Lindquist EA, Lucas S, Paulsen IT, Hattenrath-Lehmann TK, Talmage SC, Walker EA, Koch F, Burson AM, Marcoval MA, Tang Y-Z, LeCleir GR, Coyne KJ, Berg GM, Bertrand EM, Saito MA, Gladyshev VN, Grigoriev IV. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. Proc Natl Acad Sci U S A 108:4352–4357. https://doi.org/10.1073/pnas.1016106108.
5. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of Nanopore sequencing to the genomics community. Genome Biol 17:239. https://doi.org/10.1186/s13059-016-1103-0.
6. Truchon AR, Gann ER, Wilhelm SW. 2022. Extraction of high molecular weight DNA from Aureococcus anophagefferens Virus. protocols.io. https://www.protocols.io/view/extraction-of-high-molecular-weight-dna-from-aureo-n2bvjx7owlk5/v1.
7. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y.
8. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132.
9. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
10. Bushnell B. 2022. BBMap. https://sourceforge.net/projects/bbmap.
11. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
12. Dunigan D, Agarkova IV. 2016. Isolation and purification of DNA from *Chlorella* viruses. protocolsio https://doi.org/10.17504/protocols.io.eq5bdy6.
13. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
14. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
15. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides 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.
16. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
17. 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.
18. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.
19. Blanc-Mathieu R, Dahle H, Hofgaard A, Brandt D, Ban H, Kalinowski J, Ogata H, Sandaa R-A. 2021. A persistent giant algal virus, with a unique morphology, encodes an unprecedented number of genes involved in energy metabolism. J Virol 95:e02446-20. https://doi.org/10.1128/JVI.02446-20.
20. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645. https://doi.org/10.1101/gr.092759.109.