

4-13-2023

## Decontamination and Annotation of the Draft Genome Sequence of the Oomycete *Lagenidium giganteum* ARSEF 373

William R. Morgan  
*The College of Wooster*

Aurelien Tartar  
*Nova Southeastern University, aurelien@nova.edu*

Follow this and additional works at: [https://nsuworks.nova.edu/cnso\\_bio\\_facarticles](https://nsuworks.nova.edu/cnso_bio_facarticles)



Part of the [Biology Commons](#)

---

### NSUWorks Citation

Morgan, William R. and Aurelien Tartar. 2023. "Decontamination and Annotation of the Draft Genome Sequence of the Oomycete *Lagenidium giganteum* ARSEF 373." *Microbiology Resource Announcements* 12, (5). doi:10.1128/mra.01346-22.

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact [nsuworks@nova.edu](mailto:nsuworks@nova.edu).



# Decontamination and Annotation of the Draft Genome Sequence of the Oomycete *Lagenidium giganteum* ARSEF 373

 William R. Morgan,<sup>a</sup> Aurélien Tartar<sup>b</sup>

<sup>a</sup>Department of Biology, The College of Wooster, Wooster, Ohio, USA

<sup>b</sup>Department of Biological Sciences, Nova Southeastern University, Fort Lauderdale, Florida, USA

**ABSTRACT** Scaffolds of a previously published *Lagenidium giganteum* ARSEF 373 genome assembly found at GenBank were filtered to remove contaminating sequences. Genome annotation of the 437 scaffolds (total length, 56.2 MB; GC content, 58.8%) with confirmed *L. giganteum* sequences identified 13,069 potential protein-coding genes, encoding at least 737 predicted secreted proteins and >100 putative translocated effectors.

The oomycete *Lagenidium giganteum* is recognized as a biocontrol agent for mosquitoes, especially *Aedes aegypti* (1). With a goal of developing environmentally sustainable compounds for controlling mosquito populations, genomic studies have been undertaken to better understand the molecular mechanism by which *L. giganteum* infects mosquito larvae (2, 3).

An unannotated *L. giganteum* ARSEF 373 whole-genome assembly (accession number [GCA\\_002286825.1](https://www.ncbi.nlm.nih.gov/nuclink/GCA_002286825.1)) is available in the NCBI Genome database (4), but we found that both *cox1* gene sequences in this assembly differ extensively (~92% identity) from *cox1* sequences previously amplified from *L. giganteum* isolates (5). Furthermore, one was nearly identical (>99.9%) to a DNA sequence ([SPLM01000079.1](https://www.ncbi.nlm.nih.gov/nuclink/SPLM01000079.1)) in the oomycete *Pythium oligandrum*, a close relative of *Lagenidium* (6), suggesting the presence of contaminating DNA sequences.

To explore the extent of the contamination, we aligned the scaffold sequences (GenBank: [NSDO00000000.1](https://www.ncbi.nlm.nih.gov/nuclink/NSDO00000000.1)) using NCBI BLASTN v2.11.0 with default parameters to collections of *Pythium oligandrum* (NCBI txid41045) and *L. giganteum* (txid4803) DNA and mRNA sequences. Of the 2,779 scaffolds, 689 extensively matched *P. oligandrum* sequences (>99% identity, >1,000 bp), indicating extensive contamination. Of interest here, 437 scaffolds (56.2 Mb; GC content, 54.84%;  $N_{50}$ , 202 kb) extensively match *L. giganteum* sequences.

The 437 scaffolds with confirmed *L. giganteum* sequences were annotated using Maker2 v3.01.03 (7, 8). The annotation evidence included 34,733 *L. giganteum* PacBio transcriptome sequencing (RNA-seq) reads (mean length, 1,122.6 bp), 608,747 *Oomycota* (NCBI txid4762) mRNA sequences retrieved from NCBI's nucleotide database (accessed June 2021; mean length, 830.8 bp), and 83 *L. giganteum* (txid4803) and 1,303,914 *Oomycota* (txid4762) protein sequences retrieved from NCBI's Identical Protein Groups database (accessed June 2021; mean lengths, 268.8 amino acids [aa] and 423.1aa, respectively).

After repeat sequences were masked using an *L. giganteum*-specific repeat library built with RepeatModeler v2.0.2a (9), protein-coding gene predictions were inferred in the first round of Maker using mRNA alignments and protein homology (8). After training Augustus v3.3.3 (10) and SNAP v2019-06-03 (11) with the first-round gene predictions, the optimized parameters were used in MAKER for a round of *ab initio* gene prediction. Finally, putative gene functions were assigned based on protein homology to UniProt/Swiss-Prot entries (12). After quality filtering, the final build contained 13,069 annotated genes.

Several measures indicate a good quality genome annotation. First, the genome size and number of predicted genes is comparable to those of related genomes (13, 14). Second, 96% of gene annotations had a low (<0.5) annotation edit distance (15), and 54% had a

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

**Copyright** © 2023 Morgan and Tartar. 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 William R. Morgan, [wmorgan@wooster.edu](mailto:wmorgan@wooster.edu).

The authors declare no conflict of interest.

**Received** 23 January 2023

**Accepted** 23 March 2023

**Published** 13 April 2023

recognizable protein domain, exceeding MAKER's "rule-of-thumb" values (90% and 50%, respectively) for a well-annotated genome (8). Furthermore, analysis using BUSCO v5.2.2 (16) found that the *L. giganteum* protein set contained complete copies of 92% of 100 stramenopile and 57.4% of 758 fungal "core" orthologs, comparable to genome annotation projects in related species (17). Comparable to *Pythium* genomes, the *L. giganteum* genome include 737 genes predicted to encode secreted proteins (18), with over 100 similar to RXLR or Crinkler pathogen effectors (19–22). This annotated genome will enhance future studies investigating the mechanisms of *L. giganteum* pathogenesis and support the development of more sustainable approaches for controlling mosquito populations.

**Data availability.** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [DAKRPA000000000](https://doi.org/10.1093/nar/gk1290) and the BioProject accession number [PRJNA256125](https://doi.org/10.1093/nar/gk1290). The version described in this paper is [DAKRPA010000000](https://doi.org/10.1093/nar/gk1290). The raw reads were deposited in the SRA under accession number [SRR21186961](https://doi.org/10.1093/nar/gk1290). The repeat library and other supporting files are available at <https://doi.org/10.6084/m9.figshare.c.6466708>.

## ACKNOWLEDGMENTS

This research was supported in part by a grant from the U.S. Department of Agriculture (Agriculture and Food Research Initiative 2011-68004-30104) and the Theron L. Peterson and Dorothy R. Peterson Biology Research Fund (College of Wooster). Technical support for the PacBio IsoSeq sequencing was provided by the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR) NextGen DNA Sequencing Core Facility (RRID:SCR\_019152). High-performance computing resources and technical support were provided by the Ohio Supercomputer Center (Columbus, OH).

## REFERENCES

- Skovmand O, Kerwin J, Lacey LA. 2007. Microbial control of mosquitoes and black flies, p 735–750. In Lacey LA, Kaya HK (ed), *Field manual of techniques in invertebrate pathology*. Springer, Dordrecht, Netherlands.
- Quiroz Velasquez PF, Abiff SK, Fins KC, Conway QB, Salazar NC, Delgado AP, Dawes JK, Douma LG, Tartar A. 2014. Transcriptome analysis of the entomopathogenic oomycete *Lagenidium giganteum* reveals putative virulence factors. *Appl Environ Microbiol* 80:6427–6436. <https://doi.org/10.1128/AEM.02060-14>.
- Olivera IE, Fins KC, Rodriguez SA, Abiff SK, Tartar JL, Tartar A. 2016. Glycoside hydrolases family 20 (GH20) represent putative virulence factors that are shared by animal pathogenic oomycetes, but are absent in phytopathogens. *BMC Microbiol* 16:232. <https://doi.org/10.1186/s12866-016-0856-7>.
- NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 44:D7–D19. <https://doi.org/10.1093/nar/gkv1290>.
- Leoro-Garzon P, Gonedes AJ, Olivera IE, Tartar A. 2019. Oomycete metabarcoding reveals the presence of *Lagenidium* spp. in phytotelmata. *PeerJ* 7:e7903. <https://doi.org/10.7717/peerj.7903>.
- McGowan J, Fitzpatrick DA. 2020. Recent advances in oomycete genomics. *Adv Genet* 105:175–228. <https://doi.org/10.1016/bs.adgen.2020.03.001>.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>.
- Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics* 48:4.11.1–4.11.39. <https://doi.org/10.1002/0471250953.bi0411s48>.
- Smit A, Hubble R. 2008. RepeatModeler Open-1.0.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a Web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33:W465–W467. <https://doi.org/10.1093/nar/gki458>.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
- The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489. <https://doi.org/10.1093/nar/gky092>.
- Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Lerksuthirat T, Tangphatsornruang S, Suriyaphol P, Grenville-Briggs LJ, Garg G, Kittichotirat W, Krajaejun T. 2015. Draft genome sequence of the pathogenic oomycete *Pythium insidiosum* strain Pi-5, isolated from a patient with pythiosis. *Genome Announc* 3:e00574-15. <https://doi.org/10.1128/genomeA.00574-15>.
- Hathcock T, Mendoza L, Liles MR, Kang Y, Newton J, Bond N, Kalalah A, Waits DS, Wang C. 2018. Draft genome sequence of the mammalian pathogen *Lagenidium giganteum* strain MTLA-03. *J Bacteriol Mycol* 5:1075.
- Eilbeck K, Moore B, Holt C, Yandell M. 2009. Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics* 10:67. <https://doi.org/10.1186/1471-2105-10-67>.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- Kushwaha SK, Vetukuri RR, Grenville-Briggs LJ. 2017. Draft genome sequence of the mycoparasitic oomycete *Pythium oligandrum* strain CBS 530.74. *Genome Announc* 5:e00346-17. <https://doi.org/10.1128/genomeA.00346-17>.
- Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
- Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-García A, Ammar R, Staskawicz BJ, Kamoun S. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* 19:2349–2369. <https://doi.org/10.1105/tpc.107.051037>.
- Jiang RHY, Tyler BM. 2012. Mechanisms and evolution of virulence in oomycetes. *Annu Rev Phytopathol* 50:295–318. <https://doi.org/10.1146/annurev-phyto-081211-172912>.
- Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AMV, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JIB, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grünwald NJ, Horn K, Horner NR, Hu C-H, Huitema E, Jeong D-H, Jones AME, Jones JDG, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393–398. <https://doi.org/10.1038/nature08358>.
- McGowan J, Fitzpatrick DA. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. *mSphere* 2:e00408-17. <https://doi.org/10.1128/mSphere.00408-17>.