AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Draft Genome Sequence of *Capniomyces stellatus*, the Obligate Gut Fungal Symbiont of Stonefly

Yan Wang,[a,b] Merlin M. White,[c] Jean-Marc Moncalvo[a,b]

Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada[a]; Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada[b]; Department of Biological Sciences, Boise State University, Boise, Idaho, USA[c]

***Capniomyces stellatus* is a host-specific endosymbiotic fungus, living in the hindgut of stoneflies (especially in *Allocapnia*). Here, we present the first draft genome sequence of the fungus, as well as the *ab initio* gene prediction and function analyses, which will facilitate the study and comparative analyses of insect-associated fungi.**

Received 9 June 2016 **Accepted** 15 June 2016 **Published** 4 August 2016

Citation Wang Y, White MM, Moncalvo J-M. 2016. Draft genome sequence of *Capniomyces stellatus*, the obligate gut fungal symbiont of stonefly. Genome Announc 4(4): e00804-16. doi:10.1128/genomeA.00761-16.

Address correspondence to Yan Wang, yanxw.wang@mail.utoronto.ca.

The gut fungus *Capniomyces stellatus*, belonging to Harpellales (Kickxellomycotina), was originally isolated from the hindgut of winter-emerging stonefly nymph, *Allocapnia granulata* (Capniidae) (1). *Capniomyces stellatus* was found strictly associated with the stonefly species and less frequently encountered, except in Arkansas, Missouri, and Tennessee (2, 3). The narrow distribution of the fungus may be due to the inability of long-distance flight of *Allocapnia* spp. and stringent habitat requirements, like the clarity of the water, cool temperature, and special food sources (4).

Capniomyces stellatus (MIS-10-108, ARSEF 9258) was obtained from the USDA-ARS Collection of Entomopathogenic Fungal Cultures (ARSEF). Cultures were reproduced using brain heart infusion glucose tryptone medium at room temperature (3). DNA extraction followed the standard CTAB protocol (5). Four PCR-free libraries—500-bp paired-end (PE), 3-kb mate-pair (MP), 5-kb MP, and 10-kb MP—were prepared and sequenced using the Illumina HiSeq 2500 platform (2 × 125-bp read length) at the Donnelly Sequencing Centre, University of Toronto (Toronto, Canada). Raw sequence reads were adapter-trimmed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Contigs were assembled using Ray version 2.3.1 (6), and scaffolds were built using SSPACE (7). Satellites, simple repeats, and low-complexity sequences were annotated with RepeatMasker version 4.0.5 (http://www.repeatmasker.org) and Tandem Repeat Finder version 4.07b (8), corresponding to fungal sequences from RepBase (9). Gene prediction employed AUGUSTUS version 3.1 (10) using the genome profile of *Conidiobolus coronatus* (Entomophthoramycotina, Zygomycota) (11). Gene functions were annotated as previously described (12), using Blast2GO version 3.2 and InterproScan version 5.8-49.0 (13, 14). TransDecoder (15) was used to predict open reading frames and enable a conservative comparison to estimate gene numbers. CEGMA version 2.4.010312 was used to identify the presence of core eukaryotic protein-coding genes and subsequent evaluation of genome coverage (16). Secreted proteins were predicted using SignalP version 4.1 (17), and transmembrane helices were predicted through the TMHMM server version 2.0 (18).

The presented genome is based on the assembled libraries sequencing data, which amounts to 426 Mb (PE), 1.1 Gb (3-kb MP), 511 Mb (5-kb MP), and 1.3 Gb (10-kb MP), respectively, representing a 130-fold genome coverage on average. The assembly consists of 72 scaffolds, with a total size of 24.8 Mb (N_{50}, 596 kb). The GC content is 37.8%. We identified 241 out of the 248 eukaryotic core genes. The *ab initio* gene prediction discovered 7,181 genes containing 15,865 exons, as well as 7,831 open reading frames in total. A total of 4.54% of the assembly was identified as low complexity or simple or interspersed repeats, which was masked using RepeatMasker. Functional annotation resulted in gene ontology terms for 3,968 genes and InterPro domains for 5,673 genes, 871 (12.1% of the total) of which were predicted as secreted proteins.

Nucleotide sequence accession numbers. This whole-genome project has been deposited in DDBJ/ENA/GenBank under the accession number LUVW00000000. The version described in this paper is the first version, LUVW01000000.

ACKNOWLEDGMENTS

We thank R. Humber and K. Hansen for preparing and providing the fungal strains. We also thank the SciNet staff at the University of Toronto for facilitating access to the supercomputing infrastructure.

This work was supported by the University of Toronto Fellowship to Y.W. and by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant 453847 to J-M.M. M.M.W. gratefully acknowledges support from the ZyGoLife project, National Science Foundation (NSF) award DEB1441715, for ongoing research.

REFERENCES

1. **Peterson SW, Lichtwardt RW.** 1983. *Capniomyces stellatus* and *Simuliomyces spica*: new taxa of Harpellales (Trichomycetes) from winter-emerging stoneflies. Mycologia **75:**242–250. http://dx.doi.org/10.2307/3792808.
2. **Lichtwardt RW, Huss MJ, Williams MC.** 1993. Biogeographic studies on trichomycete gut fungi in winter stonefly nymphs of the genus *Allocapnia*. Mycologia **85:**535–546. http://dx.doi.org/10.2307/3760499.

3. **White MM, Siri A, Lichtwardt RW**. 2006. Trichomycete insect symbionts in Great Smoky Mountains National Park and vicinity. Mycologia **98:** 333–352. http://dx.doi.org/10.3852/mycologia.98.2.333.
4. **Peterson SW, Lichtwardt RW**. 1987. Antigenic variation within and between populations of three genera of Harpellales (Trichomycetes). Trans Br Mycol Soc **88:**189–197. http://dx.doi.org/10.1016/S0007-1536(87)80214-0.
5. **Wang Y, Tretter ED, Johnson EM, Kandel P, Lichtwardt RW, Novak SJ, Smith JF, White MM**. 2014. Using a five-gene phylogeny to test morphology-based hypotheses of *Smittium* and allies, endosymbiotic gut fungi (Harpellales) associated with arthropods. Mol Phylogenet Evol **79:** 23–41. http://dx.doi.org/10.1016/j.ympev.2014.05.008.
6. **Boisvert S, Laviolette F, Corbeil J**. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol **17:**1519–1533. http://dx.doi.org/10.1089/cmb.2009.0238.
7. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W**. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:** 578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
8. **Benson G**. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27:**573–578. http://dx.doi.org/10.1093/nar/27.2.573.
9. **Bao W, Kojima KK, Kohany O**. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA **6:**11. http://dx.doi.org/10.1186/s13100-015-0041-9.
10. **Keller O, Kollmar M, Stanke M, Waack S**. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics **27:**757–763. http://dx.doi.org/10.1093/bioinformatics/btr010.
11. **Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, LaButti KM, Lindquist EA, Yee–Ngan C, Ohm RA, Salamov AA, Grigoriev IV, Spatafora JW, Berbee ML**. 2015. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. Genome Biol Evol **7:**1590–1601. http://dx.doi.org/10.1093/gbe/evv090.
12. **Wang Y, White MM, Kvist S, Moncalvo J-M**. 24 June 2016. Genome-wide survey of gut fungi (Harpellales) reveals the first horizontally transferred ubiquitin gene from a mosquito host. Mol Biol Evol. [Epub ahead of print.] http://dx.doi.org/10.1093/molbev/msw126.
13. **Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M**. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research Bioinformatics **21:**3674–3676. http://dx.doi.org/10.1093/bioinformatics/bti610.
14. **Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S**. 2014. InterProScan 5: genome-scale protein function classification. BioInformatics **30:**1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031.
15. **Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A**. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc **8:**1494–1512. http://dx.doi.org/10.1038/nprot.2013.084.
16. **Parra G, Bradnam K, Korf I**. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23:**1061–1067. http://dx.doi.org/10.1093/bioinformatics/btm071.
17. **Petersen TN, Brunak S, von Heijne G, Nielsen H**. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **8:**785–786. http://dx.doi.org/10.1038/nmeth.1701.
18. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL**. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567–580. http://dx.doi.org/10.1006/jmbi.2000.4315.