AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Metagenome-Assembled Genomes of Bacteria Associated with *Massospora cicadina* Fungal Plugs from Infected Brood VIII Periodical Cicadas

Cassandra L. Ettinger,[a,b] Brian Lovett,[c] Matt T. Kasson,[c] Jason E. Stajich[a,b]

[a]Department of Microbiology and Plant Pathology, University of California—Riverside, Riverside, California, USA
[b]Institute for Integrative Genome Biology, University of California—Riverside, Riverside, California, USA
[c]Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia, USA

ABSTRACT We report six metagenome-assembled genomes (MAGs) associated with *Massospora cicadina* strain MCPNR19 (ARSEF 14555), an obligate entomopathogenic fungus of periodical cicadas. The MAGs include representatives of *Pantoea*, *Pseudomonas*, *Lactococcus*, and one potential new *Chryseobacterium* species. Future research is needed to resolve the ecology of these MAGs and determine whether they represent symbionts or contaminants.

Massospora cicadina (Zoopagomycota) is an understudied obligate entomopathogenic fungus that infects periodical cicadas (1, 2). During assembly and annotation of an improved genomic resource for *M. cicadina* strain MCPNR19 (ARSEF 14555) (3) using BlobTools2 (4), we identified many bacterial contigs. Given the discovery of psychoactive compounds present in *Massospora* species (5) and the uncertainty regarding their production, coupled with the established relationships between cicadas and bacterial mutualists (6), we sought to bin metagenome-assembled genomes (MAGs) from the improved *M. cicadina* genome to inform future investigations into tripartite cicada-*Massospora*-bacteria interactions.

The sampling, extraction, sequencing, quality control, and assembly methods are described by Stajich et al. (3). Briefly, conidia and azygospores of *M. cicadina* MCPNR19 (ARSEF 14555) were collected from fungal plugs of infected brood XIII Pharaoh cicadas (*Magicicada septendecim*) in June 2019. Genomic DNA from spores was sequenced using both Illumina and Nanopore technologies. The Nanopore data were assembled using wtdbg2 v. 2.5 (7), followed by multiple rounds of polishing with the Illumina reads (8).

We screened the preliminary *M. cicadina* genome assembly from Stajich et al. (3) for MAGs using the Anvi'o v. 7 pipeline (9). First, we calculated the genomic coverage against the Illumina reads using Bowtie2 v. 2.4.2 (10) and SAMtools v. 1.11 (11). We then used "anvi-gen-contigs-database" to generate a database from the *M. cicadina* genome assembly and called open-reading frames on this database using Prodigal v. 2.6.3 (12). As part of the Anvi'o pipeline, we identified bacterial (13), archaeal (13), and protista (14) single-copy genes using HMMER v. 3.2.1 (15) and rRNA genes using Barrnap (16). A predicted taxonomy was assigned to each gene call using Kaiju v. 1.7.2 (17) with the NCBI BLAST nonredundant (nr) protein database v. 2020-05-25, which included fungi and microbial eukaryotes. An Anvi'o profile was then constructed using "anvi-profile" for contigs >2.5 kbp with the "–cluster-contigs" option. The automatic binning algorithm, MetaBAT2 v. 2.12.1 (18), was run on contigs >2.5 kbp from the *M. cicadina* genome assembly to generate the preliminary MAGs. These MAGs were imported into Anvi'o using "anvi-import-collection" and were then manually inspected, combined, and refined using "anvi-interactive" and "anvi-refine." The completeness and redundancy of the MAGs was

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.
For a companion article on this topic, see https://doi.org/10.1128/MRA.00367-22.
The authors declare no conflict of interest.
Received 25 April 2022
Accepted 18 May 2022
Published 29 August 2022

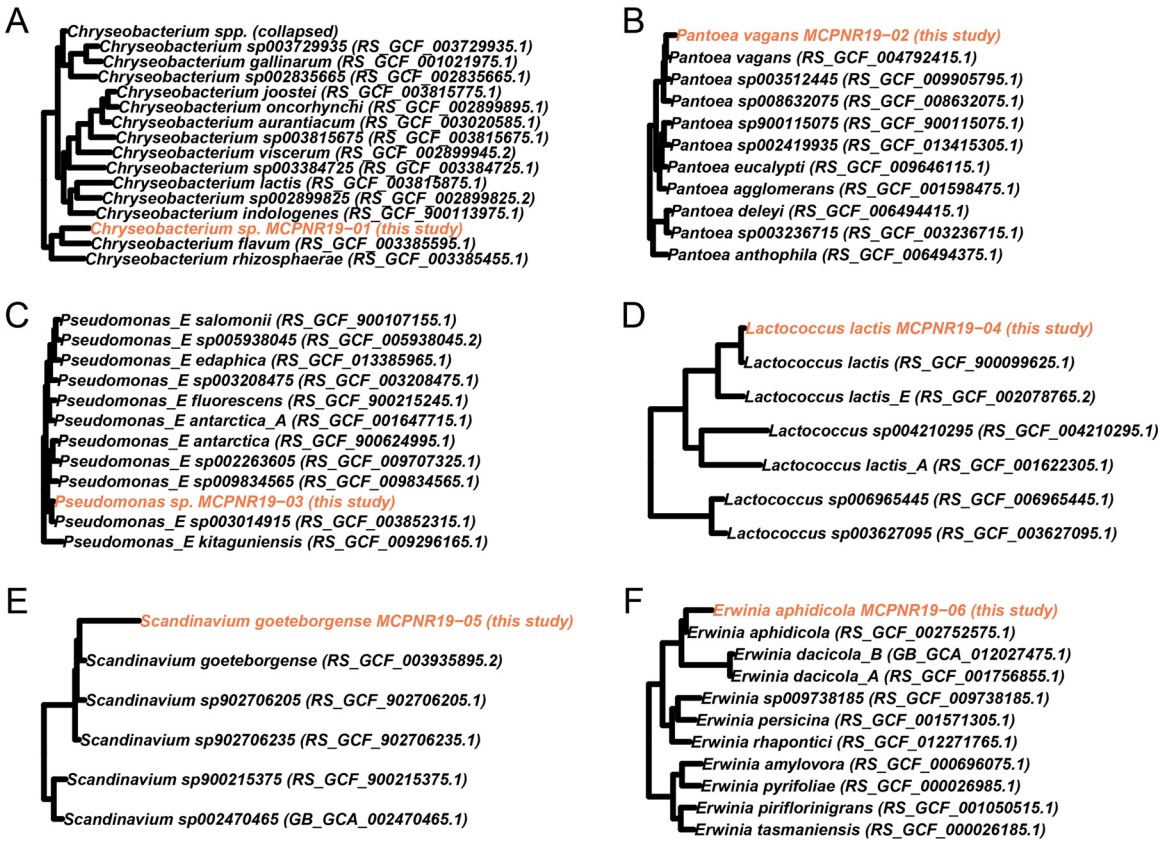

FIG 1 Phylogenetic placement of draft MAGs using GTDB-Tk. Phylogenetic trees were produced using GTDB-Tk for (A) *Chryseobacterium* sp. MCPNR19-01, (B) *Pantoea vagans* MCPNR19-02, (C) *Pseudomonas* sp. strain MCPNR19-03, (D) *Lactococcus lactis* MCPNR19-04, (E) *Scandinavium goeteborgense* MCPNR19-05, and (F) *Erwinia aphidicola* MCPNR19-06. MAGs generated in this study are highlighted in orange.

assessed within Anvi'o using "anvi-summarize" and then again using the CheckM v. 1.1.3 lineage-specific workflow (19). We defined the MAGs as high quality if they were >90% complete, medium quality if >50% complete, and low quality if <50% complete as in reference 20. To obtain a putative taxonomy for each MAG, we used GTDB-Tk v. 1.5.0 (21), which places MAGs phylogenetically in the context of the Genome Taxonomy Database. We visualized the MAG phylogenetic placement in R v. 4.1.2 using the ggtree v. 3.2.1 and Treeio v. 1.18.1 packages (22–24).

We generated a total of six draft MAGs (two high-quality and single-contig, two medium-quality, and two low-quality MAGs) representing six genera, including a high-quality MAG for a potentially new *Chryseobacterium* species (Fig. 1; Table 1). No MAGs were obtained representing the known cicada bacterial mutualists *Hodgkinia* or *Sulcia*. The higher genomic coverage of MAGs in the azygospore sequencing libraries may indicate that these MAGs represent opportunistic infections of moribund cicadas. Ultimately, these MAGs will provide valuable future insight into *Massospora*-bacterial interactions and symbiosis.

Data availability. The sequence reads have been deposited under SRA accession numbers SRR17553520 to SRR17553526 and BioProject accession number PRJNA795459. The four medium- and high-quality MAG assemblies have been deposited at DDBJ/ENA/GenBank under accession numbers JALIHL000000000, JALIHM000000000, JALIHN000000000, and JALIHO000000000. The versions described in this paper are JALIHL010000000, JALIHM010000000, JALIHN010000000, and JALIHO010000000. The two low-quality MAG assemblies are archived at Zenodo (25). Related computational scripts for this work are available on GitHub and archived in Zenodo (26).

TABLE 1 Summary of genomic features for metagenome-assembled genomes identified in association with *Massospora cicadina*

Bin identifier	Draft quality[a]	Putative taxonomy	Closest taxonomic placement ANI (%)[b]	Genome size (bp)	No. of contigs	N_{50} (bp)	GC content (%)	Illumina mean coverage (×)[c]	Nanopore mean coverage (×)[d]		No. of genes[e]	CheckM % completion	CheckM % redundancy	Anvi'o % completion	Anvi'o % redundancy	GenBank accession no. or reference
									Azygospores	Conidia						
MCPNR19-01	High	*Chryseobacterium* sp.	82.96	4,875,107	1	4,875,107	37.60	16.11	25.62	2.46	4,647	100	0.61	97.18	2.82	JALIHL010000000
MCPNR19-02	High	*Pantoea vagans*	97.36	3,966,629	1	3,966,629	55.49	43.32	56.22	0.89	3,824	93.42	0.08	98.59	2.82	JALIHM010000000
MCPNR19-03	Medium	*Pseudomonas* sp.	95.96	6,113,041	63	666,019	60.52	316.80	102.70	0.03	5,730	87.25	2.49	95.78	0	JALIHN010000000
MCPNR19-04	Medium	*Lactococcus lactis*	98.38	2,214,970	5	892,863	34.95	5.22	28.33	0.99	2,560	85.29	1.7	94.37	2.82	JALIHO010000000
MCPNR19-05	Low	*Scandinavium goeteborgense*	95.47	1,822,081	68	39,073	55.75	21.35	19.65	1.42	2,118	45.61	0	35.21	0	25
MCPNR19-06	Low	*Erwinia aphidicola*	97.49	1,786,331	62	35,544	56.90	22.98	15.83	6.16	1,991	29.31	0	30.99	1.41	25

[a] Draft quality was assigned based on CheckM completion and redundancy estimates.
[b] Average nucleotide identity (ANI) to closest taxonomic placement in the Genome Taxonomy Database as reported by GTDB-Tk.
[c] Illumina data sets were generated from fungal azygospores.
[d] Nanopore data sets were generated from either fungal azyogospores or conidia.
[e] Number of genes predicted using Prodigal.

ACKNOWLEDGMENTS

J.E.S. is a CIFAR fellow in the program "Fungal Kingdom: Threats and Opportunities" and was supported by the U.S. Department of Agriculture National Institute of Food and Agriculture Hatch project CA-R-PPA-211-5062-H and National Science Foundation (NSF) award DEB-1441715. Illumina library preparation was completed at the University of California-Riverside in the Institute of Integrative Genome Biology (IIGB), and the libraries were sequenced at the UC Berkeley Vincent J. Coates facility. Metagenome-assembled genome assembly was performed at the IIGB High-Performance Computing Cluster, supported by NSF grant DBI-1429826 and NIH grant S10-OD016290.

We thank John Wenzel (CMNH) for permission to sample the cicadas, Amy Metheny and Angie Macias for assistance with collection, and Matthew Collins and Holly Clark (UCR, IIGB) for library preparation.

REFERENCES

1. Lovett B, Macias A, Stajich JE, Cooley J, Eilenberg J, de Fine Licht HH, Kasson MT. 2020. Behavioral betrayal: how select fungal parasites enlist living insects to do their bidding. PLoS Pathog 16:e1008598. https://doi.org/10.1371/journal.ppat.1008598.
2. Macias AM, Geiser DM, Stajich JE, Łukasik P, Veloso C, Bublitz DC, Berger MC, Boyce GR, Hodge K, Kasson MT. 2020. Evolutionary relationships among Massospora spp. (Entomophthorales), obligate pathogens of cicadas. Mycologia 112:1060–1074. https://doi.org/10.1080/00275514.2020.1742033.
3. Stajich JE, Lovett BR, Ettinger CL, Carter-House DA, Kurbessoian T, Kasson MT. 2022. An improved 1.5-gigabase draft assembly of *Massospora cicadina* (Zoopagomycota), an obligate fungal parasite of 13- and 17-year cicadas. Microbiol Resour Announc. https://doi.org/10.1128/mra.00367-22.
4. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit—interactive quality assessment of genome assemblies. G3 (Bethesda) 10:1361–1374. https://doi.org/10.1534/g3.119.400908.
5. Boyce GR, Gluck-Thaler E, Slot JC, Stajich JE, Davis WJ, James TY, Cooley JR, Panaccione DG, Eilenberg J, De Fine Licht HH, Macias AM, Berger MC, Wickert KL, Stauder CM, Spahr EJ, Maust MD, Metheny AM, Simon C, Kritsky G, Hodge KT, Humber RA, Gullion T, Short DPG, Kijimoto T, Mozgai D, Arguedas N, Kasson MT. 2019. Psychoactive plant- and mushroom-associated alkaloids from two behavior modifying cicada pathogens. Fungal Ecol 41:147–164. https://doi.org/10.1016/j.funeco.2019.06.002.
6. Simon C, Cooley JR, Karban R, Sota T. 2022. Advances in the evolution and ecology of 13- and 17-year periodical cicadas. Annu Rev Entomol 67:457–482. https://doi.org/10.1146/annurev-ento-072121-061108.
7. Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods 17:155–158. https://doi.org/10.1038/s41592-019-0669-3.
8. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
9. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.
10. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
12. 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.
13. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162–4164. https://doi.org/10.1093/bioinformatics/btz188.
14. Delmont T. 5 May 2018. Assessing the completion of eukaryotic bins with anvi'o. Meren Lab. https://merenlab.org/2018/05/05/eukaryotic-single-copy-core-genes/.
15. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
16. Seemann T. 2013. Barrnap: basic rapid ribosomal RNA predictor. https://github.com/tseemann/barrnap.
17. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7:11257. https://doi.org/10.1038/ncomms11257.
18. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359.
19. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
20. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, The Genome Standards Consortium, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.
21. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
22. Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T, Zhu H, Guan Y, Jiang Y, Yu G. 2020. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Mol Biol Evol 37:599–603. https://doi.org/10.1093/molbev/msz240.
23. Yu G. 2020. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinformatics 69:e96. https://doi.org/10.1002/cpbi.96.
24. R Core Team. 2021. R: a language and environment for statistical computing. https://www.R-project.org/.
25. Ettinger CL, Lovett BR, Kasson MT, Stajich JE. 2022. Metagenome-assembled genomes of *Scandinavium goeteborgense* MCPNR19-05 and *Erwinia aphidicola* MCPNR19-06. Zenodo. https://doi.org/10.5281/zenodo.6426149.
26. Ettinger CL, Stajich JE. 2022. stajichlab/Massospora_cicadina_MAGs (v.1). Zenodo. https://doi.org/10.5281/zenodo.6455718.