Draft Genome Sequence of the Yeast *Kodamaea ohmeri*, a Symbiont of the Small Hive Beetle

James P. Tauber,ᵃ Anna K. Childers,ᵃ Jay D. Evansᵃ

ᵃU.S. Department of Agriculture, Bee Research Laboratory, Beltsville Agricultural Research Center, Beltsville, Maryland, USA

ABSTRACT *Kodamaea ohmeri* is a symbiont of the small hive beetle (SHB), which is a scavenger of honey bee colonies. The SHB causes absconding of the economically important honey bee (*Apis mellifera*) and deposits *K. ohmeri* in the honeycomb. We describe long-read sequencing and further analyses of the *K. ohmeri* genome.

The small hive beetle (SHB), *Aethina tumida*, deposits its symbiotic yeast (*Kodamaea ohmeri*, Saccharomycotina) when infesting honey bee (*Apis mellifera*) colonies (1, 2). *K. ohmeri* ferments honey, creating an odor and slimy appearance used for SHB diagnostics (3), as well as volatiles that mimic honey bee pheromones, which may attract more SHBs to the hive (2). *K. ohmeri* is also a human pathogen (4) and produces ethanol and food flavors (5).

Macerated midguts of bees from a SHB-infested honey bee colony were plated on yeast-peptone-dextrose agar (Sigma) with antibiotics and incubated at 25°C. A single colony was picked to produce an overnight culture, which was pelleted for genomic DNA (gDNA) extraction using the MagAttract high-molecular-weight (HMW) DNA kit (Qiagen). Primers ITS1_F, ITS4_R, NL1_F, and NL4_R were used in Phusion (NEB) reactions for identification.

A whole-genome sequence (WGS) library was constructed using the SMRTbell express template prep kit v1, size selected on a BluePippin system with a 15-kb cutoff, and sequenced on a PacBio Sequel system. One single-molecule real-time (SMRT) cell produced 6,057,833,952 bp in 835,386 reads for >475× coverage. PacBio data were corrected, trimmed, and assembled into 27 contigs with Canu v1.7.1 (6) using the parameters genomeSize = 12.16m and correctedErrorRate = 0.105. One contig flagged as circular was identified as the mitochondrion by BLAST. MUMmer v4.0.0beta2 (7) confirmed the overlapping ends. Circularization and removal of contigs contained within another contig were performed with Circlator v1.5.0 (−b2r_length_cutoff 60000 −split_all_reads [8]). Three iterations of Arrow (SMRT Link v5.1.0) polishing produced a 20-contig assembly (QUAST [9] and Galaxy v4.6.3 [10]), with 12,583,843 bp, 42.75% GC content, an N_{75} value of 1,584,252 bp, an L_{75} value of 5, and a largest contig of 2,525,569 bp. BUSCO v3 (11) assessment using the Fungi odb9 and Saccharomycetales odb9 data sets indicated 99% and 93.8% completeness, respectively. MAKER 3.01.02-beta (12) with AUGUSTUS v3.3.2 (13) and trained *Candida guilliermondii*, the previous species before reassignment and closest relative available (14), predicted 5,239 genes. Gene descriptions were assigned via BLASTP (BLAST+ 2.8 [15]) against the Reviewed (Swiss-Prot) database following support protocol 3 (16). BlastKOALA v2.1 (17), eggNOG-mapper (DIAMOND v4.5.1) (18), and GO FEAT v1.0 (19) assigned 58.9% proteins with KEGG orthology identifiers and 4,574 proteins with Clusters of Orthologous Groups. Twenty-eight percent of the annotatable predicted proteins are involved in cellular processes/signaling, 25% in information storage/processing, and 28% in metabolism. antiSMASH v5-beta (fungi) (20) identified a squalene synthase but no prominent natural product gene clusters for secondary metabolism. Because the SHB genome possesses

no invertase for sucrose metabolism (21), we searched the proteome of *K. ohmeri* for genes that might assist in gut metabolism and identified an invertase and a hexose transporter. Identification of transaldolase and xylulose kinase genes corroborates the production of ethanol by xylose-fermenting *K. ohmeri* strains (5). To identify *K. ohmeri* sequences in the SHB genome assembly as speculated based on BLASTN, the *K. ohmeri* assembly was fragmented (500 bp) using BBMap v38 (22) and mapped to the SHB genome (assembly number GCA_001937115) using HISAT 2.1.0 (23), followed by SAMtools and BLASTN. Ribosomal DNA genes were mainly identified among the 1.14% of the *K. ohmeri* assembly that aligned to SHB (0.64% once and 0.50% more than once) using BLASTN. NCBI's SRA taxonomy analysis found 1.6% alignment.

This genome will help unravel bee-SHB-yeast tripartite interactions, improve next-generation sequencing SHB studies, and advance clinical and industrial efforts.

Data availability. Data were deposited at GenBank under BioProject number PRJNA525764 (BioSample number SAMN11074024, SRA number SRR8889280, and assembly number SKFK00000000). The strain was deposited at CBS-KNAW Collections–Westerdijk Fungal Biodiversity Institute as CBS 15370.

REFERENCES

1. Amos BA, Leemon DL, Hayes RA, Cribb BW, Furlong MJ. 2018. Associations between the small hive beetle and the yeast Kodamaea ohmeri throughout the host life cycle. J Econ Entomol 111:1501–1508. https://doi.org/10.1093/jee/toy121.
2. Torto B, Boucias DG, Arbogast RT, Tumlinson JH, Teal P. 2007. Multitrophic interaction facilitates parasite–host relationship between an invasive beetle and the honey bee. Proc Natl Acad Sci U S A 104: 8374–8378. https://doi.org/10.1073/pnas.0702813104.
3. Cornelissen B, Neumann P. 2018. How to catch a small beetle: top tips for visually screening honey bee colonies for small hive beetles. Bee World 95:99–102. https://doi.org/10.1080/0005772X.2018.1465374.
4. Al-Sweih N, Khan ZU, Ahmad S, Devarajan L, Khan S, Joseph L, Chandy R. 2011. Kodamaea ohmeri as an emerging pathogen: a case report and review of the literature. Med Mycol 49:766–770. https://doi.org/10.3109/13693786.2011.572300.
5. Sharma S, Arora A, Sharma P, Singh S, Nain L, Paul D. 2018. Notable mixed substrate fermentation by native Kodamaea ohmeri strains isolated from Lagenaria siceraria flowers and ethanol production on paddy straw hydrolysates. Chem Cent J 12:8. https://doi.org/10.1186/s13065-018-0375-8.
6. 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.
7. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
8. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
9. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
10. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

12. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2007. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196. https://doi.org/10.1101/gr.6743907.

13. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–312. https://doi.org/10.1093/nar/gkh379.

14. Brandt ME, Lockhart SR. 2012. Recent taxonomic developments with Candida and other opportunistic yeasts. Curr Fungal Infect Rep 6:170–177. https://doi.org/10.1007/s12281-012-0094-x.

15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

16. 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.

17. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.

18. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.

19. Araujo FA, Barh D, Silva A, Guimarães L, Ramos R. 2018. GO FEAT: a rapid Web-based functional annotation tool for genomic and transcriptomic data. Sci Rep 8:1794. https://doi.org/10.1038/s41598-018-20211-9.

20. Blin K, Pascal Andreu V, de los Santos ELC, Del Carratore F, Lee SY, Medema MH, Weber T. 2019. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 47:D625–D630. https://doi.org/10.1093/nar/gky1060.

21. Evans JD, McKenna D, Scully E, Cook SC, Dainat B, Egekwu N, Grubbs N, Lopez D, Lorenzen MD, Reyna SM, Rinkevich FD, Neumann P, Huang Q. 2018. Genome of the small hive beetle (Aethina tumida, Coleoptera: Nitidulidae), a worldwide parasite of social bee colonies, provides insights into detoxification and herbivory. GigaScience 7:giy138. https://doi.org/10.1093/gigascience/giy138.

22. Bushnell B. 2016. BBMap short-read aligner, and other bioinformatics tools. https://sourceforge.net/projects/bbmap/.

23. 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.