Whole-Genome Sequence of the *Wolbachia* Strain *w*Tcon, an Endosymbiont of the Confused Flour Beetle, *Tribolium confusum*

Yeganeh Gharabigloozare,ᵃ Alexander Wähling,ᵃ,ᵇ Christoph Bleidornᵃ

ᵃDepartment of Animal Evolution and Biodiversity, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, University of Göttingen, Göttingen, Germany
ᵇInstitute of Computer Science, Campus Institute of Data Science, University of Göttingen, Göttingen, Germany

ABSTRACT Up to 60% of insects are infected with symbiont intracellular alphaproteobacteria of the genus *Wolbachia*, which are often able to manipulate their host's reproduction. Here, we report the annotated draft genome sequence of strain *w*Tcon from the confused flour beetle, *Tribolium confusum*, based on long- and short-read sequence data. The assembled genome is located on 12 contigs with a total size of 1,418,452 bp.

Tenebrionid beetles of the genus *Tribolium* include cosmopolitan pests of many stored products, from flour to dried foods (1). Due to minimal maintenance work and access to all life stages, two different tenebrionid beetle species, the red flour beetle (*Tribolium castaneum*) and the confused flour beetle (*Tribolium confusum*), have been studied in the lab for decades as model systems (2). *Tribolium confusum* is known to be naturally infected with *Wolbachia* (3–5), an intracellular alphaproteobacterium which has been found as a widespread endosymbiont of arthropods and nematodes (6). To support the utility of *Wolbachia* in the *Tribolium* model system, we report the draft genome sequence of *w*Tcon.

DNA was extracted from adult *T. confusum* beetles of the *Wolbachia*-infected strain MN61 (originally collected in Kansas, USA), supplied by the Stored Product Insect and Engineering Research Unit of USDA-ARS. Beetles were reared on flour medium and brewer's yeast (5%) at 30°C and 65% relative humidity with a 16:8-h dark/light cycle. For Oxford Nanopore Technologies (ONT) long reads, high-molecular-weight (HMW) DNA from beetles (10 males and 10 females) was extracted using both a MagAttract HMW DNA kit (Qiagen, Hilden, Germany) and phenol-chloroform extraction with precipitation by sodium acetate (7). Two ONT libraries were prepared (SQK-LSK109)—without fragmentation or size selection—and sequenced using MinION (FLO-MIN 106) protocols from ONT. The raw ONT reads from two separate runs were base called using Guppy v3.4.4, yielding 4,074,131 reads with an N_{50} value of 2,511 bp. Furthermore, quality assessment and trimming were performed using FastQC v0.11.9 (8) and NanoFilt v2.8.0 (9), respectively. For short-read sequencing, genomic DNA was isolated using the Quick-DNA Miniprep Plus kit (Zymo Research, Irvine, CA, USA). Library preparation and sequencing were performed by AllGenetics (A Coruña, Spain), using Illumina paired-end (150-bp) libraries, sequenced using an Illumina HiSeq 2500 instrument, yielding 46,330,553 paired-end reads, which were trimmed using FastQC v0.11.8 (8). Default parameters were used except where otherwise noted.

The initial metagenome assembly for the ONT long reads was generated using Flye v2.8.1 (10), which estimated the size of the genome at around 1.4 Mbp. Using a BLAST search against the reference genome of the supergroup B *Wolbachia* strain *w*PipPel from *Culex quinquefasciatus* (GenBank accession number NC_010981.1), we identified putative *Wolbachia* contigs in the metagenome, and by mapping reads using Minimap2 v2.13-r850 (11) onto these contigs, we retrieved putative *Wolbachia* ONT raw reads. For Illumina reads, we created a metagenome assembly using SPAdes v3.13.2 with a *k*-mer of 77 (12). After

mapping the trimmed raw reads onto this assembly using Bowtie v2.3.4.1 (13), we were able to extract short reads of putative *Wolbachia* origin.

The hybrid assembly of *w*Tcon using the extracted long and short reads was completed using Unicycler v0.4.9 (14), which contains 12 contigs (N_{50} value, 138,551 bp) in a total length of 1,418,452 bp (GC content, 34.1%). The draft genome was annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (15). The annotation identified 1,236 protein-coding genes, 58 pseudogenes, 34 tRNAs, 4 noncoding RNAs (ncRNAs), and 3 rRNAs (5S, 16S, and 23S). We assessed the genome completeness using BUSCO v5.2.2 (16). Out of 364 searched BUSCO groups, 353 were complete and single copy, 2 complete and duplicated, 11 missing, and none fragmented, resulting in a 96.9% BUSCO completeness score. We compared the five housekeeping markers of the multilocus strain typing (MLST) system (17), namely, *gatB*, *coxA*, *hcpA*, *fbpA*, and *ftsZ*, with the public PubMLST database (https://pubmlst.org/) and could confirm that our sequenced strain shows the exact profile as available for *Tribolium confusum* (strain Tcon_B_BhAvill AK; *Wolbachia* sequence type 30 [ST30]).

Data availability. The accession number for the complete genome sequence of *w*Con in GenBank is JAIZNT000000000. In addition, the raw sequence reads for Oxford Nanopore and Illumina sequencing have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA767570.

REFERENCES

1. Campbell JF, Arthur FH, Mullen MA. 2004. Insect management in food processing facilities. Adv Food Nutr Res 48:239–295. https://doi.org/10.1016/S1043-4526(04)48005-X.
2. Pointer MD, Gage MJ, Spurgin LG. 2021. *Tribolium* beetles as a model system in evolution and ecology. Heredity (Edinb) 126:869–883. https://doi.org/10.1038/s41437-021-00420-1.
3. Fialho RF, Stevens L. 1996. *Wolbachia* infections in the flour beetle *Tribolium confusum*: evidence for a common incompatibility type across strains. J Invertebr Pathol 67:195–197. https://doi.org/10.1006/jipa.1996.0032.
4. Goodacre SL, Fricke C, Martin OY. 2015. A screen for bacterial endosymbionts in the model organisms *Tribolium castaneum*, *T. confusum*, *Callosobruchus maculatus*, and related species. Insect Sci 22:165–177. https://doi.org/10.1111/1744-7917.12096.
5. Kageyama D, Narita S, Imamura T, Miyanoshita A. 2010. Detection and identification of 351 *Wolbachia* endosymbionts from laboratory stocks of stored-product insect pests and their 352 parasitoids. J Stored Products Res 46:13–19. https://doi.org/10.1016/j.jspr.2009.07.003.
6. Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat Rev Microbiol 6:741–751. https://doi.org/10.1038/nrmicro1969.
7. Montero-Mendieta S. 2018. DNA extraction using phenol-chloroform. protocols.io https://doi.org/10.17504/protocols.io.q2cdyaw.
8. Andrews S. 2013. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.
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. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17:1103–1110. https://doi.org/10.1038/s41592-020-00971-x.
11. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
12. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads, p 158–170. *In* Deng M, Jiang R, Sun F, Zhang X (ed), Research in computational molecular biology. RECOMB 2013. Lecture notes in computer science, vol 7821. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-642-37195-0_13.
13. Langmead B, Wilks C, Antonescu V, Charles R. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 35:421–432. https://doi.org/10.1093/bioinformatics/bty648.
14. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
15. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
16. 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.
17. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MC, Tettelin H, Werren JH. 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol 72:7098–7110. https://doi.org/10.1128/AEM.00731-06.