AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Host-Microbial Interactions | Announcement

Complete genome sequence of *Wolbachia* strain *w*Mel colonizing *Drosophila melanogaster* JW18 cells

Emily M. Layton,[1] **Macy T. Vance,**[1] **Richard W. Hardy,**[1] **Irene L. G. Newton**[1]

AUTHOR AFFILIATION See affiliation list on p. 4.

ABSTRACT *Wolbachia* are widespread insect endosymbionts known for manipulating host reproduction, aiding vector-borne disease control, and influencing host evolution. Here, we provide the complete genome sequence of the *Wolbachia* strain *w*Mel derived from the *Drosophila melanogaster* JW18 cell culture and assembled using a hybrid approach combining Illumina and Oxford Nanopore reads.

KEYWORDS cell culture, *Wolbachia*, symbiosis, gene loss

The *Drosophila melanogaster* JW18 cell line (obtained from Dr. Bill Sullivan [UCSC]) colonized with *w*Mel *Wolbachia* is a widely used model in the field for which there is no complete genome. We sought to identify genetic variations that may have been introduced due to long-term adaptation to cell culture.

JW18 cells were maintained in Schneider's insect medium supplemented with 10% fetal bovine serum and 1% antibiotic-antimycotic solution (penicillin-streptomycin-Gibco Amphotericin B) in unvented culture flasks kept in the dark at room temperature. Cells were scraped from a confluent flask and divided into two aliquots: one for the *Wolbachia* isolation (to increase the proportion of *Wolbachia*-derived reads) and the other for the whole-cell lysate (to preserve longer reads). The *Wolbachia* isolation used a previously published protocol (1) with minor modifications: samples were vortexed in 2 s pulses and spun at 20,000 *g* for 10 min. Both the isolated *Wolbachia* and whole-cell lysate pellets were resuspended in DNA/RNA Shield and kept at room temperature for 30 min before high-molecular weight DNA extraction using the Zymo Quick-DNA HMW MagBead Kit following the manufacturer's instructions for cells and solids. No pellet was observed after the initial microbial lysis centrifugation step. No shearing or size selection was performed.

The isolated *Wolbachia* sample was sequenced using the Nextera Small Genome DNA Library protocol (LIB-146b) on Illumina NextSeq 2000 with a P2 flow cell (100 cycles, v3 chemistry), generating 918 Mb across 7,522,085 reads (mean length: 61 bp). The whole-cell lysate was sequenced using the Oxford Nanopore Technologies Rapid Barcoding Kit 24 V14 on a MinION Mk1C with an R9.4 flow cell (#FAX48436) and base-called with Dorado v0.5.3, yielding 466,278 reads (N50: 12,413 bp, mean read length: 5,875 bp). The genome was assembled using a hybrid approach, starting with a long-read assembly, followed by short-read polishing (2) with default parameters, unless otherwise specified.

Short-read quality control was performed using fastp v0.24.0. For long-read preprocessing, adaptors were trimmed with Porechop v0.2.4 (3), and reads were filtered with Filtlong v0.2.4 (4) (--min_length 1000, --keep_percent 95). Reads were then mapped to a *D. melanogaster* reference genome (GCA_000001215.4) using minimap2 v2.26-r1175. Unmapped reads totaling 146 Mb across 24,939 reads (N50: 12,415 bp, mean read length: 5,870 bp) were assembled using Autocycler v0.1.2 (5) with no manual intervention. Reads were subsampled into six sets (11,372 reads per subset, ~52.6× depth per subset) and

Editor Elinne Becket, California State University San Marcos, San Marcos, California, USA

Address correspondence to Irene L. G. Newton, irnewton@iu.edu.

The authors declare no conflict of interest.

Received 3 July 2025
Accepted 20 October 2025
Published 10 November 2025

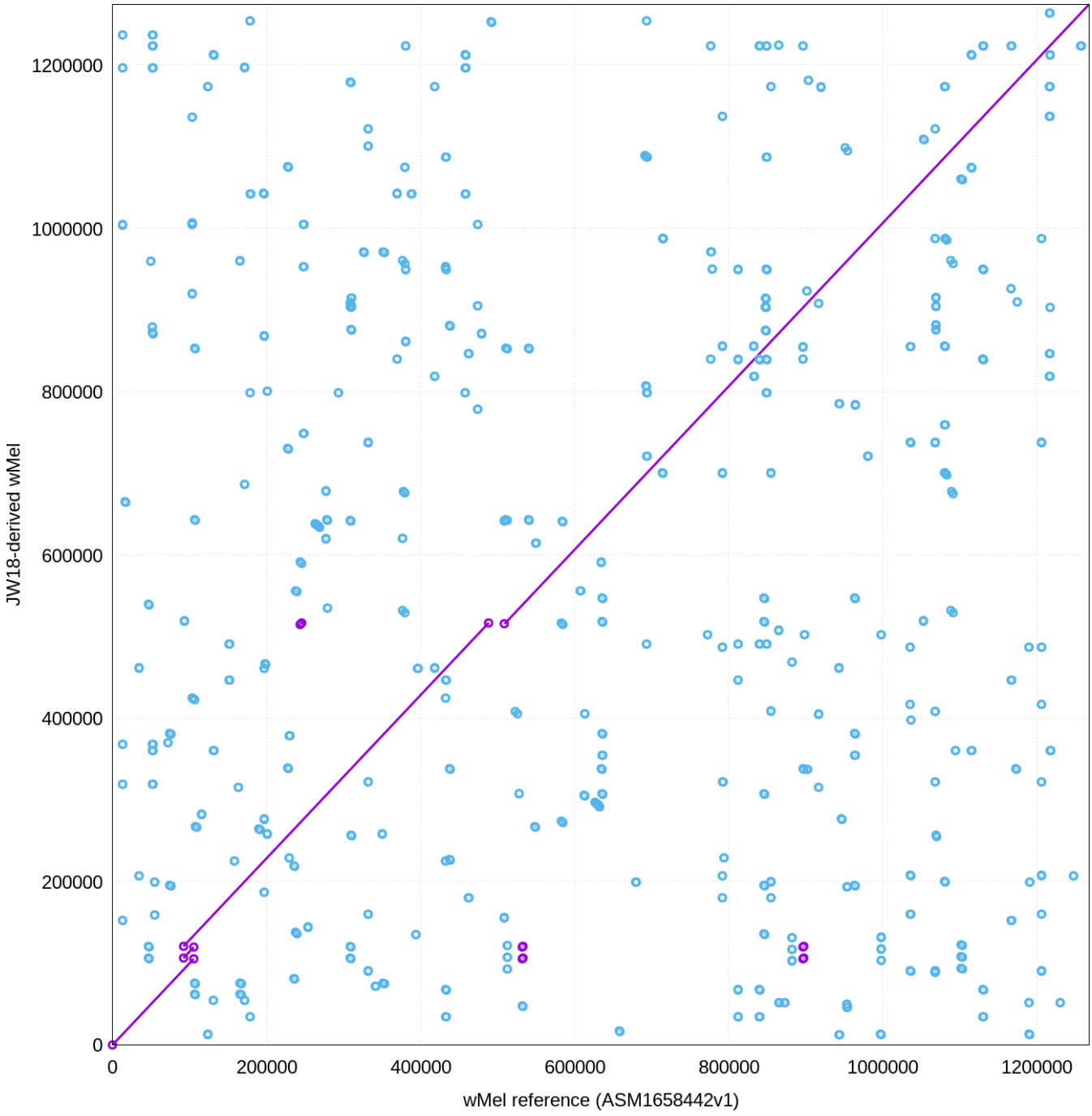

FIG 1 Alignment of the JW18-derived *w*Mel genome with the *w*Mel reference sequence. A NUCmer plot was generated using MUMmer v4.0.1 to visualize genomic rearrangements. Purple lines represent maximal unique matches between the genomes, while blue lines indicate reverse complement matches, with a minimum substring length of 20 bases. Discontinuous lines highlight a triplication and a deletion in the JW18-derived *w*Mel genome relative to the reference.

each assembled using Canu v2.2, Flye v2.9.5, miniasm v0.3, NECAT v0.0.1, NextDenovo v2.5.2, and raven v1.8.3 for a total of 36 assemblies. Contigs were clustered, and one cluster passed QC, containing a single >1.2 Mb contig from each assembler. The final consensus assembly produced a circular chromosome (1,274,542 bp, 35.2% GC, ~110× long-read depth). Autocycler detected and trimmed start-end overlaps via self-vs-self overlap alignment of the unitigs. The genome was rotated to the canonical *dnaA* start site with Dnaapler (6–10). The assembly was polished with long reads using medaka

TABLE 1 Predicted SNP and structural variation in wMel colonizing JW18 cells[a]

Position	Reads	Score	Freq	Annotation	Gene	Product
= 73311	664 + GAAGGAAG	5/80	72.7%	Coding (405/816 nt)	NEHKPIKF_00076	Hypothetical protein
724915 =				Coding (93/822 nt)	NEHKPIKF_00757	3′–5′ exonuclease
= 321592	17	13/76	9.5%	Intergenic (–66/+706)	NEHKPIKF_00341/NEHKPIKF_00342	Hypothetical protein/hypothetical protein
321794 =				Intergenic (–268/+504)	NEHKPIKF_00341/NEHKPIKF_00342	Hypothetical protein/hypothetical protein
394,635	30	N/A (SNP)	7.50%	D48G (GAC→GGC)	NEHKPIKF_00416	Bcr/CflA family efflux MFS transporter
= 395079	19	10/92	10.6%	Coding (587/1164 nt)	NEHKPIKF_00416	Bcr/CflA family efflux MFS transporter
395477 =				Coding (985/1164 nt)	NEHKPIKF_00416	Bcr/CflA family efflux MFS transporter
766095 =	45 + 24 bp	4/50	24.1%	Coding (+1280/–497)	NEHKPIKF_00795/NEHKPIKF_00796	50S ribosomal protein L9/tRNA-Met
1183286 =				Coding (1142/1230 nt)	NEHKPIKF_01227	Lipoprotein-releasing ABC transporter permease subunit
1178868 =	31	19/92	10.5%	Coding (553/582 nt)	algA_1	Alginate biosynthesis protein AlgA
1191235 =				Coding (1605/2307 nt)	NEHKPIKF_01234	AAA domain-containing protein
= 1179781	17	14/96	6.4%	Coding (28/774 nt)	algA_2	Alginate biosynthesis protein AlgA
= 1191239				Coding (1609/2307 nt)	NEHKPIKF_01234	AAA domain-containing protein

[a]SNP and structural variants were predicted by *breseq* based on the analysis of short reads relative to the JW18-derived *w*Mel consensus assembly and manually curated to exclude poorly supported predictions and extraneous information. New [EL1] junctions are predicted by split reads where each side of the read maps to a different location in the JW18-derived *w*Mel reference genome described here. Paired rows correspond to predicted new junctions; each side of the split read is represented in one row. The "Position" column lists the coordinates in the reference genome that are joined together in the junction, and the equals sign informs how the reads supporting the junction are oriented relative to the reference genome. If each side of the split read is mapped to the reference sequence, "= position" means that starting from the listed position coordinate, the read extends to lower coordinates in the reference genome, whereas "position =" means that starting from the listed position coordinate, the read extends to higher coordinates in the reference genome. The most common type of junction, for example, "= 5" and "11 =", likely indicates a deletion of bases 6–10. The single row indicates an SNP, with "Position" marking the mutation site. "Reads" specifies the number of supporting reads. "Score" denotes the position-hash (left) and minimum-overlap (right) scores. "Freq" indicates the predicted frequency of that SNP or junction in the population. "Annotation" describes the position of the SNP or junction breakpoint within a gene (coding) or upstream (+) or downstream (–) of the nearest neighboring genes (intergenic). "Gene" lists the affected or flanking genes; "Product" describes their function. See *breseq* output documentation for more details.

v2.0.1 (11) (--bacteria) and with short reads using Polypolish v0.6.0 (12, 13) and Pypolca v0.3.1 (--careful) (13). Annotation with the NCBI Prokaryotic Genome Annotation Pipeline v6.10 identified 1,243 coding sequences, 34 tRNAs, and one copy of each of the 5S, 16S, and 23S rRNA genes. Genome completeness was assessed with CheckM v1.2.4 (99.79% complete, 2.56% contamination, o_Rickettsiales (UID3811) data set) (10, 14–16) and BUSCO v6.0.0 (99.1% complete, rickettsiales odb12 data set) (17).

We identified a ~19.5 kbp deletion corresponding to the previously described Octomom region (18, 19), along with an amplification of a ~14.3 kbp region (threefold increase) containing 12 genes, including those related to metabolite synthesis and export, protein export, a DNA polymerase III subunit, and a DNA gyrase subunit, similar but not identical to a duplication previously found in JW18-derived *w*Mel (20) (Fig. 1).

Structural variation analysis using *breseq* v0.39.0 (21, 22) (--polymorphism-prediction) with both long (--nanopore) and short reads revealed several structural variants (Table 1), raising the intriguing possibility that there are subpopulations of *Wolbachia* with distinct genomic architectures that could underlie differences in phenotypes, consistent with *Wolbachia's* previously described highly recombining, mosaic genome structure (23).

ACKNOWLEDGMENTS

We gratefully acknowledge Ben Fulton and Robert Henschel for the supercomputing support. We thank Chris Robinson for assistance with Nanopore sequencing and Carrie Ganote, Patrick Foy, and Lílian Caesar for their bioinformatics guidance. We also thank members of the Newton and Hardy labs for their valuable feedback on this work. Special

thanks to Dr. William Sullivan for generously providing the *D. melanogaster* JW18-dox and JW18-*w*Mel cell cultures. Illumina genome sequencing was performed by the Center for Genomics and Bioinformatics at Indiana University.

This research was supported by the National Science Foundation under grant number EF 2025389 (to I.L.G.N. and R.W.H.), by the National Science Foundation Graduate Research Fellowship (to E.M.L.), and by an Advanced Cyberinfrastructure fellowship provided by University Information Technology Services at Indiana University (to E.M.L.). This research was also supported in part by shared university research grants from IBM, Inc. to Indiana University and by Lilly Endowment, Inc. through its support for the Indiana University Pervasive Technology Institute.

AUTHOR AFFILIATION

[1]Department of Biology, Indiana University, Bloomington, Indiana, USA

AUTHOR ORCIDs

Richard W. Hardy http://orcid.org/0000-0001-6912-6291
Irene L. G. Newton http://orcid.org/0000-0002-7118-0374

AUTHOR CONTRIBUTIONS

Macy T. Vance, Methodology.

DATA AVAILABILITY

This genome project is available at GenBank under BioProject accession number PRJNA1246252. The complete genome sequence of JW18-derived *w*Mel can be found under GenBank accession number CP186703 and the reported genome is the first version, CP186703.1. The raw reads are available from the NCBI Sequence Read Archive under the accession numbers SRR32971796 for the Illumina reads and SRR32971797 for the Oxford Nanopore FASTQ file.

REFERENCES

1. Nevalainen LB, Layton EM, Newton ILG. 2023. *Wolbachia* promotes its own uptake by host cells. Infect Immun 91:e00557–22. https://doi.org/10.1128/iai.00557-22
2. Wick RR. 2021. Choose-your-own-adventure guide to bacterial genome assembly (1.2.0). Zenodo. Available from: https://doi.org/10.5281/zenodo.7471199
3. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132
4. Wick RR. 2021. Filtlong (0.2.4). Available from: https://github.com/rrwick/Filtlong
5. Wick RR. 2025. Autocycler (0.1.2). Zenodo. Available from: https://doi.org/10.5281/zenodo.14642607
6. Bouras G, Grigson SR, Papudeshi B, Mallawaarachchi V, Roach MJ. 2024. Dnaapler: a tool to reorient circular microbial genomes. JOSS 9:5968. https://doi.org/10.21105/joss.05968
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
8. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028. https://doi.org/10.1038/nbt.3988
9. Larralde M. 2022. Pyrodigal: python bindings and interface to prodigal, an efficient method for gene prediction in prokaryotes. JOSS 7:4296. https://doi.org/10.21105/joss.04296
10. 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
11. Oxford Nanopore Technologies. 2018. medaka (2.0.1). Available from: https://github.com/nanoporetech/medaka
12. Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802
13. Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. 2024. How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. Microb Genom 10:001254. https://doi.org/10.1099/mgen.0.001254
14. 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
15. Matsen FA, Kodner RB, Armbrust EV. 2010. Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538. https://doi.org/10.1186/1471-2105-11-538
16. Eddy SR. 2011. Accelerated Profile HMM Searches. PLOS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195
17. Tegenfeldt F, Kuznetsov D, Manni M, Berkeley M, Zdobnov EM, Kriventseva EV. 2025. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res 53:D516–D522. https://doi.org/10.1093/nar/gkae987
18. Chrostek E, Marialva MSP, Esteves SS, Weinert LA, Martinez J, Jiggins FM, Teixeira L. 2013. Wolbachia variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. PLOS Genet 9:e1003896. https://doi.org/10.1371/journal.pgen.1003896
19. Chrostek E, Teixeira L. 2015. Mutualism breakdown by amplification of *Wolbachia* genes. PLOS Biol 13:e1002065. https://doi.org/10.1371/journal.pbio.1002065
20. Grobler Y, Yun CY, Kahler DJ, Bergman CM, Lee H, Oliver B, Lehmann R. 2018. Whole genome screen reveals a novel relationship between

Wolbachia levels and *Drosophila* host translation. PLoS Pathog 14:e1007445. https://doi.org/10.1371/journal.ppat.1007445
21. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165–188. https://doi.org/10.1007/978-1-4939-0554-6_12
22. Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, Knoester DB, Reba A, Meyer AG. 2014. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. BMC Genomics 15:1039. https://doi.org/10.1186/1471-2164-15-1039
23. Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SGE. 2009. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proceedings of the National Academy of Sciences 106:5725–5730. https://doi.org/10.1073/pnas.0810753106