Microbiology Resource Announcements

| Computational Biology | Announcement

Complete *de novo* assembly of *Wolbachia* endosymbiont of contemporary *Drosophila simulans* using long-read genome sequencing

Jodie Jacobs,[1,2] Alexandra Lum,[1] Elyse Mina,[1] Camryn N. Morey,[1,2] Darren D. Lee,[1] Emry Gutierrez,[1] Jonah Dionisio,[1] Cade Mirchandani,[1,2] Luke Sylvester,[1] Anne Nakamoto,[1,2] Hailey Loucks,[1,2] Ciara Wanket,[2,3] Ariana Cisneros,[1,2] Alessandro Calicchio,[1] Alexis N. Enstrom,[1] Camille Headrick,[1] Faith Okamoto,[1,2] Harrison D. Heath,[1,2] Kseniya Malukhina,[1] Petria Russell,[1,2] Sagorika Nag,[1] Thomas Gillespie,[1] William Sobolewski,[1,2] Zia Truong,[1,2] Shelbi L. Russell[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present a contemporary high-quality, complete *de novo* assembly of *Wolbachia pipientis* (*w*Ri Merrill 23, OZ411647), an alphaproteobacterial endosymbiont of *Drosophila simulans (D. simulans)*. This assembly was generated using long-read sequencing of *w*Ri-infected *D. simulans* embryos collected from Merrill College at the University of California, Santa Cruz, in October 2023.

KEYWORDS genome, symbiosis, *Wolbachia*, *Drosophila*, evolution, evolutionary biology, biological control, ecology

Wolbachia pipientis infects diverse arthropods and nematodes, manipulating host phenotypes through cytoplasmic incompatibility (CI), male killing, and fertility rescue (1, 2). The Riverside strain (*w*Ri) was first identified in California *Drosophila simulans (D. simulans)* in the 1980s (3) and rapidly spread statewide due to exceptionally strong CI (4). Despite its significance (5), the only reference genome reflects *w*Ri present in 1984 (3, 6). Here, we present a complete *de novo* *w*Ri genome assembly from contemporary *D. simulans* collected at the UC Santa Cruz Alan Chadwick Garden in Merrill College in October 2023.

To generate a contemporary *w*Ri genome assembly, we collected wild *D. simulans* flies, established isofemale lines, and sequenced *w*Ri-infected embryos. We established isofemale lines by deploying banana-baited bottles for ~5 days and collecting gravid females onto white food medium. After offspring eclosed, species identity was confirmed by phenotyping males and by PCR using silf-F/R primers to distinguish *D. simulans* from *D. melanogaster* (7) and wsp_1F/592R primers to confirm *w*Ri identity (8). We extracted DNA from *w*Ri-infected embryos using the Wizard HMW DNA Extraction Kit (Promega) without shearing and prepared libraries with the Native Barcoding Kit V14 (SQK-NBD114-24) without size selection. We sequenced these libraries on the Nanopore MinION Mk1B with an R10 flow cell (FLO-MIN-114) and MinKNOW v23.07.8 with adaptive sampling (fast model) to deplete *D. simulans* reads (GCF_016746395.2) yielding 4.26M reads, (N50 749 bp, 20 h) that were basecalled with Dorado (v0.7.3, hac model, --min-qscore 10). After filtering for host-free reads >3 kb (93K reads remaining, N50 10.5 kb), we assembled the *w*Ri genome using Flye (9) following the methods described by Jacobs and Nakamoto *et al*. (10), yielding a 1.26 Mb circular assembly with 30× coverage. Flye determined circularization (4,000 bp minimum overlap), the genome was not rotated.

To polish the assembly, we generated Illumina short-read whole-genome sequencing data from whole *w*Ri-infected *D. simulans* flies (Merrill 23 stocks). Illumina libraries were prepared using a Tn5-based tagmentation protocol (11). The Tn5 enzyme (Tn5R27S,

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Shelbi L. Russell, shelbilrussell@gmail.com.

Alessandro Calicchio, Alexis N. Enstrom, Camille Headrick, Faith Okamoto, Harrison D. Heath, Kseniya Malukhina, Petria Russell, Sagorika Nag, Thomas Gillespie, William Sobolewski, and Zia Truong contributed equally to this article. Authorship order was determined based on contributions to data generation and analysis. For authors with equal contributions, ties were resolved alphabetically by first name.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 13 September 2025
Accepted 26 February 2026
Published 30 April 2026

TABLE 1 Annotation summary statistics

wRi Merrill 23 annotation summary	
Annotation pipeline	Prokka v1.1.1
Annotation method	kingdom:bacteria
Length (bp)	1,447,516
GC Content	35.21%
Genes (total)	1,392
CDSs (total)	1,392
Genes (RNA)	37
rRNAs	1, 1, 1 (5S, 16S, 23S)
tRNAs	34
ncRNAs	0
Pseudogenes (total)	0

E54K, and L372P) was obtained from QB3 MacroLab (University of California, Berkeley). Tagmentation simultaneously fragmented genomic DNA and ligated adapter sequences using custom oligos (Tn5-A: FC-121-1030, Tn5-B: FC-121-1031, and Tn5-rev; IDT). Libraries were indexed and amplified using the Nextera XT Index Kit v2 (i5/i7 adapters, Illumina) with KAPA HiFi (Roche) and sequenced on a NovaSeqX Plus (2 × 150 bp). Reads were trimmed to remove sequencing adapters and low-quality regions using fastp v1.0.1 (11) before polishing the assembly with Pilon v1.24 (12) (849K reads). We assessed the quality

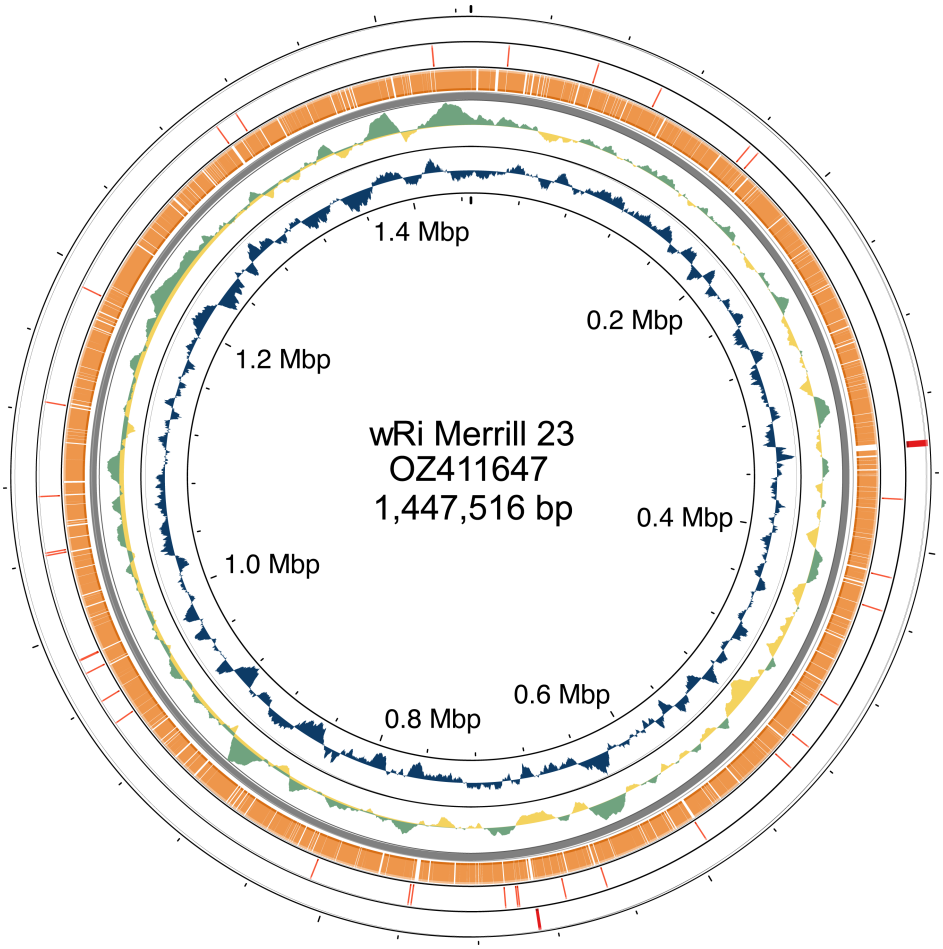

FIG 1 *Wolbachia w*Ri genome map. Concentric circles show (outer to inner) rRNA genes (red), tRNA genes (red), coding sequences (orange), GC skew (green/yellow for high/low), and GC content (blue), with GC metrics plotted as deviations from the genome-wide average.

of the polished assembly with BUSCO v5.7.0 (13) (rickettsiales_odb10), achieving 99.2% completeness; annotated the assembly with Prokka v1.1.1 (14) (kingdom:bacteria) (Table 1); and calculated and visualized GC content and GC skew with Proksee v1.1.2 (15) (Fig. 1). Default parameters were used unless otherwise specified.

ACKNOWLEDGMENTS

The authors acknowledge the University of California, Santa Cruz Genomics Institute for providing computational resources, including the Phoenix computational cluster, and support for this project.

The authors thank Rion Parsons for his support and the University of California, Santa Cruz, for use of the Hummingbird computational cluster. The authors thank James Letchinger for the use of his computer for nanopore sequencing. The authors thank the University of California, Santa Cruz Baskin Engineering Lab Support team for providing laboratory space and support.

Funding for this project was provided by National Institutes of Health (NIH) (T32 HG012344) awarded to J.J., C.W., H.L., A.N., C.S., and A.C., NIH awards to S.L.R. (R00GM135583, R35GM157189), and the NSF-GRFP awarded to A.N.

AUTHOR AFFILIATIONS

[1]Biomolecular Engineering Department, University of California, Santa Cruz, California, USA
[2]Genomics Institute at the University of California, Santa Cruz, California, USA
[3]Ecology and Evolutionary Biology Department, University of California, Santa Cruz, California, USA

AUTHOR ORCIDs

Jodie Jacobs http://orcid.org/0000-0003-0628-6268
Anne Nakamoto https://orcid.org/0000-0002-6753-7238
Shelbi L. Russell http://orcid.org/0000-0001-6734-2740

FUNDING

Funder	Grant(s)	Author(s)
National Institutes of Health	R00GM135583, R35GM157189	Shelbi L. Russell
National Institutes of Health	T32HG012344	Hailey Loucks Ciara Wanket Ariana Cisneros Faith Okamoto Harrison D. Heath
National Science Foundation		Anne Nakamoto

AUTHOR CONTRIBUTIONS

Jodie Jacobs, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing | Alexandra Lum, Data curation | Elyse Mina, Formal analysis | Camryn N. Morey, Investigation, Writing – review and editing | Darren D. Lee, Validation | Emry Gutierrez, Validation, Writing – review and editing | Jonah Dionisio, Validation | Cade Mirchandani, Data curation, Methodology, Resources | Luke Sylvester, Data curation | Anne Nakamoto, Data curation, Methodology | Hailey Loucks, Data curation | Ciara Wanket, Data curation | Ariana Cisneros, Writing – review and editing | Alessandro Calicchio, Writing – review and editing | Alexis N. Enstrom, Writing – review and editing | Camille Headrick, Writing – review and editing | Faith Okamoto, Writing – review and editing | Harrison D. Heath,

Writing – review and editing | Kseniya Malukhina, Writing – review and editing | Petria Russell, Writing – review and editing | Sagorika Nag, Writing – review and editing | Thomas Gillespie, Writing – review and editing | William Sobolewski, Writing – review and editing | Zia Truong, Writing – review and editing | Shelbi L. Russell, Conceptualization, Funding acquisition, Supervision, Writing – review and editing

DATA AVAILABILITY

Raw sequencing reads, ERX15644803 (Illumina) and ERX15644805 (ONT) and the assembled genome (OZ411647) are available under BioProject accession number PRJEB107302. Assembly pipeline: https://github.com/jodieja-cobs/Jacobs_et_al_2026_de_novo_wRi_merrill_23_assembly.

REFERENCES

1. Russell S.L, Castillo JR. 2020. Trends in symbiont-induced host cellular differentiation. Results Probl Cell Differ 69:137–176. https://doi.org/10.1007/978-3-030-51849-3_5
2. Russell SL, Castillo JR, Sullivan WT. 2023. Wolbachia endosymbionts manipulate the self-renewal and differentiation of germline stem cells to reinforce fertility of their fruit fly host. PLoS Biol 21:e3002335. https://doi.org/10.1371/journal.pbio.3002335
3. Hoffmann AA, Turelli M, Harshman LG. 1990. Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. Genetics 126:933–948. https://doi.org/10.1093/genetics/126.4.933
4. Turelli M, Hoffmann AA. 1995. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. Genetics 140:1319–1338. https://doi.org/10.1093/genetics/140.4.1319
5. Carrington LB, Lipkowitz JR, Hoffmann AA, Turelli M. 2011. A Re-examination of *Wolbachia*-induced cytoplasmic incompatibility in California *Drosophila simulans*. PLoS One 6:e22565. https://doi.org/10.1371/journal.pone.0022565
6. 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*. Proc Natl Acad Sci USA 106:5725–5730. https://doi.org/10.1073/pnas.0810753106
7. Faria VG, Sucena É. 2017. From nature to the lab: establishing *Drosophila* resources for evolutionary genetics. Front Ecol Evol 5. https://doi.org/10.3389/fevo.2017.00061
8. Casper-Lindley C, Kimura S, Saxton DS, Essaw Y, Simpson I, Tan V, Sullivan W. 2011. Rapid fluorescence-based screening for *Wolbachia* endosymbionts in *Drosophila* germ line and somatic tissues. Appl Environ Microbiol 77:4788–4794. https://doi.org/10.1128/AEM.00215-11
9. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
10. Jacobs J, Nakamoto A, Mastoras M, Loucks H, Mirchandani C, Karim L, Penunuri G, Wanket C, Russell SL. 2024. Complete *de novo* assembly of *Wolbachia* endosymbiont of *Drosophila* willistoni using long-read genome sequencing. Sci Rep 14:17770. https://doi.org/10.1038/s41598-024-68716-w
11. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560
12. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963
13. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38:4647–4654. https://doi.org/10.1093/molbev/msab199
14. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
15. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, Stothard P. 2023. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res 51:W484–W492. https://doi.org/10.1093/nar/gkad326