



Complete Genome Assemblies for Three Variants of the *Wolbachia* Endosymbiont of *Drosophila melanogaster*

 Preston J. Basting,^a  Casey M. Bergman^{a,b}

^aInstitute of Bioinformatics, University of Georgia, Athens, Georgia, USA

^bDepartment of Genetics, University of Georgia, Athens, Georgia, USA

ABSTRACT Here, we report genome assemblies for three strains of *Wolbachia pipientis*, assembled from unenriched, unfiltered long-read shotgun sequencing data of geographically distinct strains of *Drosophila melanogaster*. Our simple methodology can be applied to long-read data sets of other *Wolbachia*-infected species with limited *Wolbachia*-host lateral gene transfers to produce complete assemblies for this important model symbiont.

Wolbachia pipientis is a widespread bacterial endosymbiont that infects 40% of arthropod species (1) and induces a wide range of effects including cytoplasmic incompatibility, feminization, male killing, and parthenogenesis (2). Currently, our understanding of the impact of *Wolbachia* on its hosts is limited by the lack of complete reference genomes for different *Wolbachia* strains, with only 18 of 84 *Wolbachia* assemblies in the NCBI assembly database as of August 2019 defined as complete.

Recently, Faddeeva-Vakhrusheva et al. (3) showed that a complete assembly of *Wolbachia* could be generated as a by-product of assembling the genome of a *Wolbachia*-infected arthropod species using PacBio long-read sequences. Based on this observation, we attempted to generate complete *Wolbachia* assemblies using long-read shotgun sequencing data for three geographically distinct *Drosophila melanogaster* lines (I23 from Ithaca, NY; N25 from the Netherlands; and ZH26 from Zimbabwe) (4) that were previously identified by Early and Clark (5) as being infected with variants of the *Wolbachia* strain *wMel*. These flies were reared on a diet of 10% yeast, 10% glucose, and 1% agar at 25°C (J. Chaston, personal communication). As described by Long et al. (4), DNA was extracted by grinding ~200 adult flies in liquid nitrogen, transferring to a solution of buffer G2 with 38 μl of RNase A (100 mg/ml) and 500 μl of proteinase K (catalog number 158920; Qiagen), incubating the solution overnight at 50°C, and then extracting DNA using the Qiagen Genomic-tip kit (catalog number 10243). DNA was then sequenced on a PacBio Sequel instrument (Pacific Biosciences, Inc.) using 2 or 3 single-molecule real-time (SMRT) cells per sample. Additionally, public short-read Illumina sequencing was used for the same *D. melanogaster* lines (5). As described by Early and Clark (5), DNA for these samples was extracted from 50 adult female flies using the Qiagen DNeasy blood and tissue kit. Sequencing was performed on an Illumina HiSeq 2000 instrument to produce 100-bp paired-end reads with a 450- to 500-bp insert size. No quality control steps were applied to the PacBio or Illumina sequencing reads prior to assembly and polishing.

All reads from PacBio whole-genome shotgun sequences of each strain were assembled using CANU v1.8 (genomeSize=137.7m, useGrid=False) (6). Assemblies for each strain generated only one contig matching the *wMel* reference genome (GenBank accession number [NC_002978](https://.ncbi.nlm.nih.gov/nucl/NC_002978)) by BLASTN v2.9.0 search with default parameters (7), which in each case corresponded to the entire *Wolbachia* genome. Repetitive regions from the ends of uncircularized *Wolbachia* contigs were trimmed using minimus2 from

Citation Basting PJ, Bergman CM. 2019. Complete genome assemblies for three variants of the *Wolbachia* endosymbiont of *Drosophila melanogaster*. *Microbiol Resour Announc* 8:e00956-19. <https://doi.org/10.1128/MRA.00956-19>.

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

Copyright © 2019 Basting and Bergman. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Preston J. Basting, preston.basting@uga.edu.

Received 12 August 2019

Accepted 15 October 2019

Published 7 November 2019

TABLE 1 Accession numbers and statistics for assemblies produced and raw sequencing data used in this study

Assembly	Assembly accession no.	Genome size (bp)	GC content (%)	PacBio read accession no.	No. of PacBio reads ^a	Avg PacBio read length (bp) ^a	Illumina read accession no.	No. of Illumina reads ^a	Avg Illumina read length (bp) ^a
I23	CP042444	1,269,137	35.24	SRR7060336	31,867	4,140	SRR1663552	547,585	99
N25	CP042446	1,267,781	35.23	SRR7060337	47,126	4,096	SRR1663577	358,462	98
ZH26	CP042445	1,267,436	35.23	SRR7060339	47,960	4,407	SRR1663599	364,241	98

^a Statistics are for the subset of reads from the *Wolbachia* endosymbiont only, not the complete shotgun sequence data set.

AMOS v3.1.0 with default parameters (8). The trimmed *Wolbachia* contigs were then adjusted using BLASTN v2.9.0 (7) and faFrag (9) with default parameters so that the start of each contig matched the *wMel* reference start. The contigs were then polished using Arrow (SMRTlink v6.0.0.47841; Pacific Biosciences) and Pilon v1.23 (10) using Illumina reads from Early and Clark (5) (Table 1) mapped to the contigs using BWA-MEM v0.7.17 with default parameters (11).

After polishing, we identified 54, 13, and 18 single-nucleotide polymorphism (SNP)/indel variants for the I23, N25, and ZH26 *Wolbachia* strains, respectively, relative to the *wMel* reference genome. The higher similarity of the N25 and ZH26 *Wolbachia* strains and increased divergence of the I23 *Wolbachia* strain relative to the *wMel* reference genome are consistent with previous work showing that *Wolbachia* genomes from lines N25 and ZH26 are both in clade III of the *wMel* phylogeny (which also contains the *wMel* reference genome), while the *Wolbachia* genome from line I23 is in clade I of the *wMel* phylogeny (which is more divergent from the *wMel* reference genome) (5, 12).

Our work extends that of Faddeeva-Vakhrusheva et al. (3) by showing that high-quality, complete genome assemblies of *Wolbachia* strains can be generated without experimental enrichment of symbiont DNA (e.g., references 13 and 14). Successful *de novo* assembly of complete *Wolbachia* genomes directly from unenriched long-read sequences also demonstrates that it is unnecessary to computationally filter symbiont reads from host reads based on similarity to *Wolbachia* reference genomes prior to assembly (15, 16). We expect this process to be particularly useful for *Wolbachia*-infected hosts with few host-symbiont lateral gene transfer events, such as *D. melanogaster* (17), in which there will be few hybrid reads between host and symbiont to confound the assembly process. As the cost of long-read sequencing decreases, we argue that direct sequencing and assembly of unenriched, unfiltered long-read data sets could be applied easily to other *Wolbachia*-infected arthropod and nematode species to expand the number of complete *Wolbachia* reference genomes.

Data availability. The assemblies produced in this study were deposited at NCBI under accession number [PRJNA557362](#). PacBio data used to generate these assemblies were published by Long et al. (4) and are available under SRA accession number [SRP142531](#). Illumina data used to polish the assemblies were published by Early and Clark (5) and are available under SRA accession [SRP050151](#). Accession numbers for assemblies produced and raw read data used in this study are given in Table 1.

ACKNOWLEDGMENTS

We thank the Georgia Advanced Computing Resource Center for computational resources, Joshua Udall (Iowa State University) for providing access to raw PacBio Sequel data, and John Chaston (Brigham Young University) for information about PacBio samples used in this project.

This work was supported by a University of Georgia Research Education Award Traineeship (P.J.B.) and by the University of Georgia Research Foundation (C.M.B.).

REFERENCES

- Zug R, Hammerstein P. 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 7:e38544. <https://doi.org/10.1371/journal.pone.0038544>.
- 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>.
- Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamenone V, Suring W, Kampfraath AA, Eilers J, Le Ngoc G, van Gestel CAM, Mariën J, Smit S, van Straalen NM, Roelofs D. 2017. Coping with living in

- the soil: the genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics* 18:493. <https://doi.org/10.1186/s12864-017-3852-x>.
4. Long E, Evans C, Chaston J, Udall JA. 2018. Genomic structural variations within five continental populations of *Drosophila melanogaster*. *G3 (Bethesda)* 8:3247–3253. <https://doi.org/10.1534/g3.118.200631>.
 5. Early AM, Clark AG. 2013. Monophyly of *Wolbachia pipientis* genomes within *Drosophila melanogaster*: geographic structuring, titre variation and host effects across five populations. *Mol Ecol* 22:5765–5778. <https://doi.org/10.1111/mec.12530>.
 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. 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>.
 8. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. 2011. Next generation sequence assembly with AMOS. *Curr Protoc Bioinformatics Chapter 11:Unit 11.8*. <https://doi.org/10.1002/0471250953.bi1108s33>.
 9. Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Brief Bioinform* 14:144–161. <https://doi.org/10.1093/bib/bbs038>.
 10. Walker BJ, Abeel T, Shea T, Priest M, Abuelliel 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>.
 11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
 12. Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. *PLoS Genet* 8:e1003129. <https://doi.org/10.1371/journal.pgen.1003129>.
 13. Iturbe-Ormaetxe I, Woolfit M, Rancès E, Duploux A, O'Neill SL. 2011. A simple protocol to obtain highly pure *Wolbachia* endosymbiont DNA for genome sequencing. *J Microbiol Methods* 84:134–136. <https://doi.org/10.1016/j.mimet.2010.10.019>.
 14. Lefoulon E, Vaisman N, Frydman HM, Sun L, Foster JM, Slatko BE. 2019. Large enriched fragment targeted sequencing (LEFT-SEQ) applied to capture of *Wolbachia* genomes. *Sci Rep* 9:5939. <https://doi.org/10.1038/s41598-019-42454-w>.
 15. Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC. 2005. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6:R23. <https://doi.org/10.1186/gb-2005-6-3-r23>.
 16. Pascar J, Chandler CH. 2018. A bioinformatics approach to identifying *Wolbachia* infections in arthropods. *PeerJ* 6:e5486. <https://doi.org/10.7717/peerj.5486>.
 17. Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, Magwire MM, Blanken- Burg K, Carbone MA, Chang K, Ellis LL, Fernandez S, Han Y, High- Nam G, Hjelmén CE, Jack JR, Javaid M, Jayaseelan J, Kalra D, Lee S, Lewis L, Munidasa M, Onger F, Patel S, Perales L, Perez A, Pu L, Rollmann SM, Ruth R, Saada N, Warner C, Williams A, Wu YQ, Yamamoto A, Zhang Y, Zhu Y, Anholt RRH, Korbel JO, Mittel- Man D, Muzny DM, Gibbs RA, Barbadilla A, Johnston JS, Stone EA, Richards S, Deplancke B, Mackay T. 2014. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res* 24:1193–1208. <https://doi.org/10.1101/gr.171546.113>.