Complete Genome Sequence of *wAna*, the *Wolbachia* Endosymbiont of *Drosophila ananassae*

Mark T. Gasser,a*  Matthew Chung,a,b  Robin E. Bromley,a  Suvarna Nadendla,a  Julie C. Dunning Hotoppa,b,c

Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA

Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

Greenebaum Cancer Center, University of Maryland, Baltimore, Maryland, USA

**ABSTRACT** Here, we present the complete genome sequence of the *Wolbachia* endosymbiont *wAna*, isolated from *Drosophila ananassae* and derived from Oxford Nanopore and Illumina sequencing. We anticipate that this will aid in *Wolbachia* comparative genomics and the assembly of *D. ananassae* specifically in regions containing extensive lateral gene transfer events.

Lateral gene transfer (LGT) from the *Wolbachia* endosymbiont *wAna* in *Drosophila ananassae* constitutes >2% of the insect genome, including integrations of multiple *wAna* genomes in the abnormally large, largely heterochromatic chromosome 4 (1, 2). To aid in studies of this massive LGT, the complete *wAna* genome was obtained.

To generate an LGT-free line of *D. ananassae*, Michael Clark and John Werren at the University of Rochester introgressed *D. ananassae* harboring the *wAna* Hawaii strain into the LGT-free *D. ananassae* Florida line for 10 generations to create *D. ananassae* W2.1, which was obtained from Irene Newton at Indiana University in Bloomington. The line was reared on molasses medium in plugged bottles at 25°C and 70% relative humidity with a 12/12-h light/dark cycle. The whole flies were flash frozen in liquid nitrogen in a 50-ml Falcon tube and vortexed for 3 s, and the headless bodies were then collected with a small brush. High-molecular-weight DNA was isolated from the adult *Drosophila ananassae* W2.1 bodies using phenol-chloroform extraction with ethanol precipitation with sodium acetate (3). Illumina paired-end (2 × 150-bp) library construction and sequencing was performed using the Nextera XT library prep protocol on an Illumina MiSeq platform, yielding 51.4 Gbp in 340,594,990 sequenced reads. Long-read library preparation (SQK-RAD004) and sequencing (FLO-MIN106 R9 MinION) protocols from Oxford Nanopore Technologies (ONT) were performed with slight modifications using 2 μg of DNA as the input and omitting library-loading beads. Raw ONT read signals were base called using Albacore v2.3.1, which yielded 864.5 Mbp in 87,410 reads without barcoding or multiplexing. Sequencing adapters and possible chimeras were removed from base-called reads with Porechop v0.2.3 (4) using – discard_middle. An initial de novo assembly using only the ONT reads and miniasm v0.2.5 (5) yielded *Drosophila* and *Wolbachia* contigs. From the assembly, a single contig of the complete *wAna* genome was identified by aligning to the *wRi* genome (6) using MUMmer v3.0 (7). Illumina and ONT reads mapping to this putative *Wolbachia* contig were identified using BWA aln/sampe (8) and Minimap2 v2.10 (9) with -ax map-ont, respectively. The *Wolbachia*-mapping reads were used to construct a new, hybrid de novo assembly using Unicycler v0.4.4 (10). The assembly was visually inspected for misassemblies by remapping Illumina and ONT reads to the hybrid de novo assembly. The *wRi* genome has two nearly identical 68-kbp regions that both include a prophage (Fig. 1). We identified a 28-kbp deletion at the end of the first of these duplicated regions in the *wAna* genome (Fig. 1). This deletion was supported by ONT reads that spanned the...
deleted region but failed to assemble correctly. Therefore, the correct sequence of the first duplicate region was manually inserted after being derived from the spanning ONT reads that were Illumina corrected with Pilon v1.22 (11) and manually inspected for errors. The final, complete, and corrected assembly of the \textit{Wolbachia} endosymbiont of \textit{Drosophila ananassae}, wAna, consists of a circular chromosome of 1,401,460 bp (GC content, 35.2\%) with average sequencing depths of \(\sim 1,240 \times\) and \(\sim 12 \times\) for the Illumina and ONT reads, respectively. The genome was annotated using the IGS Prokaryotic Annotation Pipeline (12) with Prodigal v2.6.3 set to disallow calling genes that run off the edge of contigs (13). All software was run using default settings unless otherwise noted. The wAna genome contains 1,289 open reading frames (ORFs), 35 tRNA genes, and one copy of each of the 5S, 16S, and 23S rRNA genes.

**Data availability.** The complete genome sequence of wAna has been deposited in GenBank under the accession number CP042904. The Oxford Nanopore FASTQ file, Oxford Nanopore FAST5 file, and Illumina sequencing reads are available from the NCBI Sequence Read Archive (SRA) under the accession numbers SRR8306005, SRR9866440, and SRR8278850, respectively.

**ACKNOWLEDGMENTS**

This project was funded by the NSF (ABI1457957), as well as by an NIH Director’s Transformative Research Award (R01CA206188) and the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under grant number U19AI110820.

We thank John Werren and Michael Clark at the University of Rochester and Irene Newton at Indiana University for producing and sharing the fly line, respectively. Illumina genome sequencing was conducted by the Genome Resource Center at the Institute for Genome Sciences.

**REFERENCES**


