Genome Assembly of the A-Group *Wolbachia* in *Nasonia oneida* Using Linked-Reads Technology

Xiaozhu Wang [1], Xiao Xiong[1,2], Wenqi Cao[1], Chao Zhang[2], John H. Werren,[3] and Xu Wang [1,4,5,6,*]

[1]Department of Pathobiology, Auburn University

[2]Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai, China

[3]Department of Biology, University of Rochester

[4]Alabama Agricultural Experiment Station, Auburn University

[5]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama

[6]Department of Entomology and Plant Pathology, Auburn University

*Corresponding author: E-mail: xzw0070@auburn.edu.

Accepted: October 5, 2019

Data deposition: This project has been deposited at DDBJ/ENA/GenBank under the accession QESS00000000.

Abstract

Wolbachia are obligate intracellular bacteria which commonly infect various nematode and arthropod species. Genome sequences have been generated from arthropod samples following enrichment for the intracellular bacteria, and genomes have also been assembled from arthropod whole-genome sequencing projects. However, these methods remain challenging for infections that occur at low titers in hosts. Here we report the first *Wolbachia* genome assembled from host sequences using 10× Genomics linked-reads technology. The high read depth attainable by this method allows for recovery of intracellular bacteria that are at low concentrations. Based on the depth differences (714× for the insect and 59× for the bacterium), we assembled the genome of a *Wolbachia* in the parasitoid jewel wasp species *Nasonia oneida*. The final draft assembly consists of 1,293, 06 bp in 47 scaffolds with 1,114 coding genes and 97.01% genome completeness assessed by checkM. Comparisons of the five Multi Locus Sequence Typing genes revealed that the sequenced *Wolbachia* genome is the A1 strain (henceforth *w*OneA1) previously reported in *N. oneida*. Pyrosequencing confirms that the wasp strain lacks A2 and B types previously detected in this insect, which were likely lost during laboratory culturing. Assembling bacterial genomes from host genome projects can provide an effective method for sequencing bacterial genomes, even when the infections occur at low density in sampled tissues.

Key words: *Wolbachia*, *Nasonia*, parasitoid wasp, 10× Genomics Chromium linked reads, multi-locus strain typing, pyrosequencing.

Introduction

Wolbachia, alphaproteobacterial endosymbionts, are widespread and common in arthropods and filarial nematodes, either as reproductive parasites or mutualists (Werren 1997; Fenn and Blaxter 2006; Werren et al. 2008). About half of arthropods are infected with *Wolbachia* (Hilgenboecker et al. 2008; Zug and Hammerstein 2012) due to horizontal movement of the bacteria between species, although the routine mode of transmission of these bacteria is vertical through the egg cytoplasm. The jewel wasp genus of *Nasonia* has been an excellent model for *Wolbachia* research (Breeuwer and Werren 1993; Perrot-Minnot et al. 1996; Bordenstein et al. 2001, 2003; Raychoudhury et al. 2009). Eleven *Wolbachia* have so far been identified in the four species of *Nasonia* (Raychoudhury et al. 2009). These are often maintained as multiple infections within individual wasps of each species and have diverse evolutionary origins, indicating horizontal transfers from divergent host species (Raychoudhury et al. 2009). Genomic studies of *Wolbachia* blossomed in the recent years since the first complete genome of the A-*Wolbachia* parasite of *Drosophila melanogaster* published in 2004 (Wu et al. 2004). *Wolbachia* genomes are small with a range between

0.9 and 1.7 Mb. In the jewel wasp (*Nasonia*) species, only two *Wolbachia* strains have been sequenced (*w*VitA; Newton et al. 2016) and *w*VitB (Kent et al. 2011), both from *Nv*. In this study, we sequenced, assembled and annotated the *Wolbachia* strain in *Nasonia oneida (No)*, which will facilitate the comparative genomic and evolutionary analyses of this model system.

Materials and Methods

Sample Collection, DNA Extraction, and Sequencing

Genomic DNA sample was extracted from 24-h male adults of the *N. oneida* NONY strain. MagAttract DNA Mini Kit (Qiagen, MD) was used to isolate high molecular weight genomic DNA. A 10× Genomic library was constructed by using the Chromium Genome Reagent Kits v2 on 10× Chromium Controller (10× Genomics Inc., CA) and sequenced on a HiSeqX lane at the Genomic Services Lab at the HudsonAlpha Institute for Biotechnology.

Genome Assembly and Annotation

The *N. oneida* genome was assembled using the Supernova 2.1.1 assembler (Weisenfeld et al. 2017). The following steps were conducted to identify *w*OneA1 scaffolds (supplementary fig. S1, Supplementary Material online): 1) all 10× reads were aligned to the *N. oneida* assembly to calculate median coverage for each scaffold; 2) *N. oneida* scaffolds were aligned to the bacterial sequence database using BLAT v3.5 (Kent 2002) to determine percent sequence identity to known *Wolbachia* sequences; 3) assign the scaffolds to *w*OneA1 genome if they have at least 20% sequence identity with known *Wolbachia* sequences and a median coverage around 60×. The genome completeness was further evaluated by checkM (Parks et al. 2015) with default settings and Benchmarking Universal Single-Copy Orthologs (BUSCO) (Seppey et al. 2019) comparing to bacteria database. Gene annotation was conducted using DFAST prokaryotic genome annotation pipeline (Tanizawa et al. 2018). tRNA genes were predicted by tRNAscan_SE (Lowe and Eddy 1997).

Comparative Analysis of *Wolbachia* Genomes in the *Nasonia* Genus

To compare the genome structure among three sequenced *Wolbachia* genomes in *Nasonia*, we conducted whole genome alignment of *w*OneA1, *w*VitA (GCA_001983615.1) (Newton, et al. 2016), and *w*VitB (GCA_000204545.1) (Kent et al. 2011) genomes using NUCmer in the MUMmer program suite with default parameter settings (Kurtz et al. 2004). The pairwise alignments were visualized using Mummerplot (Kurtz et al. 2004). Orthologous gene sets between *w*OneA1 and two other *Wolbachia* in *Nasonia* were generated based on reciprocal best hits using BLAST with an E-value cutoff 10^{-5}. 32 genes in *w*OneA1 genome were excluded in this analysis as the gene orthology are unclear when comparing to *w*VitA and *w*VitB.

MLST Strain Typing of *Wolbachia w*OneA1 Strain

The five Multi Locus Sequence Typing (MLST) genes (Baldo 2006; Jolley and Maiden 2010) were examined to further characterize the phylogenetic relationships of *Wolbachia* strains in *Nasonia*. The *w*OneA1 MLST genes were identified on five different genome scaffolds, including *coxA* on SCAFFOLD17, *gatB* on SCAFFOLD28, *hcpA* on SCAFFOLD47, *ftsZ* on SCAFFOLD73 and *fbpA* in SCAFFOLD76. Sequences of MLST genes from the following strains were downloaded from the MLST database (Baldo 2006): *w*NvitA, *w*NvitB in *N. vitripennis; w*NgirA1, *w*NgirA2, *w*NgirB in *Nasonia giraulti*; *w*NlonA, *w*NlonBl, *w*NlonB2 in *Nasoina longicornis* (Raychoudhury et al. 2009). Multiple sequence alignments were generated using MUSCLE with default parameters (Edgar 2004). Phylogenetic analysis was performed using the Maximum Likelihood (ML) method in MEGA 7.0 software (Kumar et al. 2016). Bootstrap tests with 1,000 replicates were used to evaluate the phylogenetic trees.

Confirmation of *w*OneA1 Strain Using Pyrosequencing

Wolbachia infection types were checked in NONY, and DNA samples from a recently (July 2018) collected wild-type CAR262L strain using allele-specific pyrosequencing. Pyro PCR and sequencing primers were designed to target SNP positions in coxA and gatB genes (supplementary table S3, Supplementary Material online) in A1, A2, and B *Wolbachia* using PyroMark Assay Design 2.0 (Qiagen, USA). The A/G SNP targeted in coxA can separate B-*Wolbachia* from A1/A2-*Wolbachia*, and the C/T SNP in gatB allowed us to distinguish A1-*Wolbachia* from A2/B-*Wolbachia*. Pyrosequencing was performed on a Pyromark Q48 instrument (Qiagen, USA) using PyroMark Q48 Advanced Reagents (Qiagen, USA). Three technical replicates were performed for each sample.

Results and Discussion

Assembly of *Wolbachia* Genome *w*OneA1

This *Wolbachia* project emerged from a de novo assembly of the parasitoid wasp *No* genome (see Materials and Methods section). *Wolbachia* scaffolds were separated from the *No* genome assembly using a custom bioinformatics pipeline (supplementary figs. S1 and S2*A*, Supplementary Material online). *No* scaffolds were BLATed against bacterial genome database (Kent 2002), and we identified *Wolbachia* scaffolds based on the median coverage and sequence identity to known *Wolbachia* sequences (supplementary fig. S2, Supplementary Material online and see Materials and Methods section). We have identified this genome to be

Table 1
*w*OneA1 Assembly Summary Statistics and Comparison with *w*VitA and *w*VitB Genomes

	*w*OneA1	*w*VitA	*w*VitB
Number of contigs	65	142	509
Number of scaffolds	47	N/A	426
Contig N50 (kb)	35.88	13.38	5.79
Scaffold N50 (kb)	128.97	N/A	6.21
Number of proteins	1,114	1,042	845
Assembled genome size (bp)	1,293,406	1,211,929	1,107,643
BUSCO completeness (%)	86.5	87.2	85.1
checkM completeness (%)	97.01	99.79	99.57
checkM contamination (%)	0	0.64	1.71

from the *w*OneA1 *Wolbachia* with median genome coverage of 59.38×, which is significantly lower compared with 713.59× for the *No* genomic scaffolds (*P*-value $< 2.2 \times 10^{-16}$) and the 20,000× mitochondrial genome (supplementary fig. S2*B*, Supplementary Material online). The differences in coverage and guanine-cytosine (GC) content further assisted in the separation of the *w*OneA1 scaffolds.

The *w*OneA1 draft genome contains 1,293,406 nucleotides with 47 scaffolds and N50 of 128.97 kb. 1,114 genomes were annotated in the *w*OneA1 genome including protein coding genes, 5S, 16S, and 23S rRNA and tRNA genes. The number of contigs and scaffolds are fewer than *w*VitA and *w*VitB, and the contig and scaffold N50s are longer (table 1). The genome completeness is 97.01% accessed by and checkM, which is comparable with *w*VitA and *w*VitB, suggesting high assembly quality. *w*VitA and *w*VitB have slightly higher completeness, but at a cost of 1–2% of contamination (table 1). The BUSCO completeness is 86.5%, which is typical for complete *Wolbachia* genomes (Sinha et al. 2019).

Due to the intracellular lifestyle and inability of media culture in *Wolbachia*, the purification of *Wolbachia* DNA from the host sample can be challenging. Different methods have been applied to purify the *Wolbachia* genomic DNA from the host DNA (Klasson et al. 2009; Mavingui et al. 2012; Duplouy et al. 2013; Ellegaard et al. 2013; Brelsfoard et al. 2014; Newton et al. 2016; Badawi et al. 2018). For the *w*VitB genome project, a high-density tiled oligonucleotide array was developed to enrich for *Wolbachia* gDNA (Kent et al. 2011). An alternative approach has been to extract *Wolbachia* reads from the host whole genome sequence data set, and then align to the reference genome of the closely related *Wolbachia* strains, or perform de novo assembly using the filtered reads (Darby et al. 2012; Saha et al. 2012; Siozios et al. 2013; Lindsey et al. 2016; Chung et al. 2017).

If no prior knowledge is available about the presence of specific microbes, sequencing without purification is preferred to identify other intracellular symbionts, as well as characterizing bacterial species in insect gut microbiota at the whole-genome level. In our study, we perform de novo assembly of the host genome and *Wolbachia* genome using the 10× Genomics linked-reads technology. The *Wolbachia* DNA fragments were labeled with unique 10× barcodes, therefore they are much less likely to be misassembled into the host genome scaffolds. The microbe (*w*OneA1) and the host (*No*) have a 12-fold difference in coverage (714× vs 59×). Using most sequencing technologies, it would be difficult to assemble the bacterial genome because of insufficient coverage against the host genome. However, the 10× sequencing linked read technology has permitted accurate identification of the bacterial scaffolds, despite the relatively low abundance of the bacteria DNA in the insect. This finding is similar to a recent study showing the efficacy of 10× technology in assembling high-quality microbial genome drafts in microbiome samples (Bishara et al. 2018). Therefore, the *Wolbachia* genome assembled with 10× linked reads was of good quality with no contamination of host nuclear and mitochondrial DNA. As the cost of PacBio sequencing decreases, the long-read platforms would be better for symbionts genome assembly when the bacteria can be enriched in the sample, or the infection occurs at a high level. However, microbial genome assembly by 10× sequencing technology will likely continue to have an advantage for some time in cases where microbial associates occur at low levels in hosts or tissue samples.

Comparative Genomic Analysis of *Wolbachia* Strains in *Nasonia* Species

When comparing the gene contents of these *Wolbachia* strains, a total of 645 genes were shared among genomes of *w*OneA1, *w*VitA and *w*VitB; 212 more genes were shared between the *w*OneA1 and *w*VitA genomes but not with *w*VitB genome (fig. 1*D*). Among the 210 *w*OneA1-specific genes, a large fraction belongs to hypothetical protein ($N = 173$) and transposon-related ($N = 22$) genes. Regarding the insertion elements (IS), the *w*OneA1 genome contains similar numbers of IS elements when compared with the genomes of *w*VitA and *w*VitB (supplementary table S1, Supplementary Material online). Although *w*VitA and *w*VitB infect the same host (*Nasonia vitripennis*) and *w*OneA1 infect a different host (*N. oneida*), the gene content of *w*VitA is closer to that of *w*OneA1 than *w*VitB, as expected by their supergroup affiliations and indicating that there is no rampant recombination between the *w*VitA and *w*VitB at genome-wide level. Taken together, the results indicate that A and B *Wolbachia* retain their genetic differences even when they infect the same host, which suggests that recombination among them is not common, with the exception of phage related genes (Bordenstein and Wernegreen 2004).

Absence of A2 and B *Wolbachia* in the Assembled *N. oneida* Strain

The whole genome alignments between *w*OneA1 and *w*VitA (fig. 1*A* and supplementary table S2, Supplementary Material

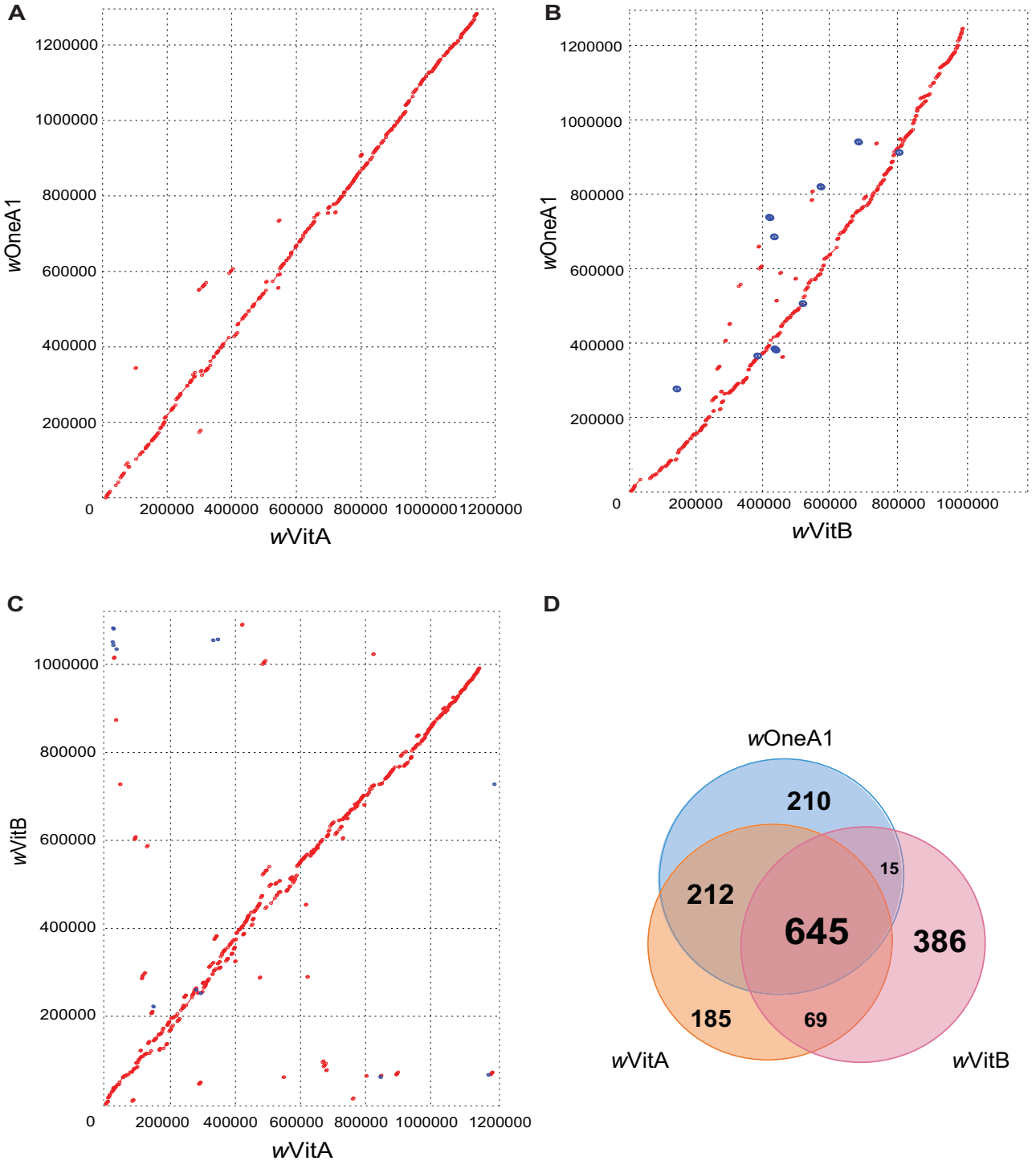

Fig. 1.—Comparative genomic analysis of *w*OneA1, *w*VitA, and *w*VitB genomes. (*A*) Dot plot showing comparison between *w*OneA1 and *w*VitA genomes, red for a forward match and blue for a reverse match; (*B*) Dot plot showing comparison between *w*OneA1 and *w*VitB genomes; (*C*) Dot plot showing comparison between *w*VitA and *w*VitB genomes; (*D*) Venn Diagram showing comparison of genes and pseudogenes in *w*OneA1, *w*VitA and *w*VitB.

online) and the phylogenetic analysis of MLST genes (supplementary fig. S3, Supplementary Material online) indicated the identified strain in our study is in the A supergroup. Furthermore, strain typing of *Wolbachia* was performed on *No* of our study and *No* genomic DNA samples that are known to be infected with all three strains (A1, A2, and B),

using independent allele-specific pyrosequencing approach (supplementary table S3, Supplementary Material online). An A/G SNP in the coxA gene was used to separate B-*Wolbachia* from A-*Wolbachia* (A allele in A1/A2-*Wolbachia* and G allele in B-*Wolbachia*, supplementary fig. S4*A*, Supplementary Material online). In gatB gene, a C/T SNP can distinguish A1-*Wolbachia* allele from A2/B-*Wolbachia* (supplementary fig. S4*B*, Supplementary Material online). The pyrosequencing results confirmed the lack of A2 and B strains in the genome assembled NONY strain. All three *Wolbachia* infections (A1, A2, and B) were successfully identified in the CAR262L strain DNA samples (supplementary fig. S4, Supplementary Material online). Therefore, the allele-specific pyrosequencing validation experiments confirmed the absence of A2 and B-type *Wolbachia* infections in the lab *No* strain (NONY). In *No* DNA samples from a recently collected field strain (CAR262L), we estimate that A1 is the dominate strain and accounts for 55% of the total infection, 40% of the infection came from the B strain and only 5% from A2 strain (supplementary fig. S4, Supplementary Material online). The absence of A2 and B *Wolbachia* in the lab *No* strain is likely due to stochastic loss during laboratory maintenance and diapause.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This project is supported by an Auburn University Intramural Grant Program Award to X.W. (AUIGP180271) and a generous laboratory startup fund to X.W. from Auburn University College of Veterinary Medicine. X.W. is also supported by Alabama Agricultural Experiment Station Enabling Grant. This work is supported by the USDA National Institute of Food and Agriculture, Hatch project 1018100. X.X. and W.C. are supported by Auburn University Presidential Graduate Research Fellowship. Contributions of J.H.W. were supported by US National Science Foundation IOS 1456233 and the Nathaniel and Helen Wisch Professorship. We thank HudsonAlpha Discovery for Illumina sequencing and Ting Li for help with DNA extractions.

Literature Cited

Badawi M, Moumen B, Giraud I, Greve P, Cordaux R. 2018. Investigating the molecular genetic basis of cytoplasmic sex determination caused by Wolbachia endosymbionts in terrestrial isopods. Genes (Basel) 9.

Baldo L. 2006. Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Appl Environ Microbiol. 72:7098–7110.

Bishara A, et al. 2018. High-quality genome sequences of uncultured microbes by assembly of read clouds. Nat Biotechnol. 36(11):1067.

Bordenstein SR, O'Hara FP, Werren JH. 2001. Wolbachia-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature 409(6821):707–710.

Bordenstein SR, Uy JJ, Werren JH. 2003. Host genotype determines cytoplasmic incompatibility type in the haplodiploid genus *Nasonia*. Genetics 164(1):223–233.

Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol. 21(10):1981–1991.

Breeuwer JA, Werren JH. 1993. Cytoplasmic incompatibility and bacterial density in *Nasonia vitripennis*. Genetics 135(2):565–574.

Brelsfoard C, et al. 2014. Presence of extensive Wolbachia symbiont insertions discovered in the genome of its host Glossina morsitans morsitans. PLoS Negl Trop Dis. 8(4):e2728.

Chung M, Small ST, Serre D, Zimmerman PA, Dunning Hotopp JC. 2017. Draft genome sequence of the Wolbachia endosymbiont of *Wuchereria bancrofti* wWb. Pathog Dis. 75.

Darby AC, et al. 2012. Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 22(12):2467–2477.

Duplouy A, et al. 2013. Draft genome sequence of the male-killing Wolbachia strain wBol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomics 14(1):20.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE. 2013. Comparative genomics of Wolbachia and the bacterial species concept. PLoS Genet. 9(4):e1003381.

Fenn K, Blaxter M. 2006. Wolbachia genomes: revealing the biology of parasitism and mutualism. Trends Parasitol. 22(2):60–65.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. 2008. How many species are infected with Wolbachia? A statistical analysis of current data. FEMS Microbiol Lett. 281(2):215–220.

Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11(1):595.

Kent BN, et al. 2011. Complete bacteriophage transfer in a bacterial endosymbiont (Wolbachia) determined by targeted genome capture. Genome Biol Evol. 3:209–218.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12(4):656–664.

Klasson L, et al. 2009. The mosaic genome structure of the Wolbachia wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci U S A. 106(14):5725–5730.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.

Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5(2):R12.

Lindsey AR, Werren JH, Richards S, Stouthamer R. 2016. Comparative genomics of a parthenogenesis-inducing wolbachia symbiont. G3 (Bethesda) 6:2113–2123.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.

Mavingui P, et al. 2012. Whole-genome sequence of Wolbachia strain wAlbB, an endosymbiont of tiger mosquito vector *Aedes albopictus*. J Bacteriol. 194(7):1840.

Newton IL, et al. 2016. Comparative genomics of two closely related Wolbachia with different reproductive effects on hosts. Genome Biol Evol. 8(5):1526–1542.

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(7):1043–1055.

Perrot-Minnot MJ, Guo LR, Werren JH. 1996. Single and double infections with Wolbachia in the parasitic wasp *Nasonia vitripennis*: effects on compatibility. Genetics 143(2):961–972.

Raychoudhury R, Baldo L, Oliveira D, Werren JH. 2009. Modes of acquisition of wolbachia: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. Evolution 63(1):165–183.

Saha S, et al. 2012. Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a Wolbachia wDi draft genome. PLoS One 7(11):e50067.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 1962: 227–245.

Sinha A, Li Z, Sun L, Carlow CKS. 2019. Complete genome sequence of the Wolbachia wAlbB endosymbiont of *Aedes albopictus*. Genome Biol Evol. 11(3):706–720.

Siozios S, et al. 2013. draft genome sequence of the Wolbachia endosymbiont of Drosophila suzukii. Genome Announc. 1(1):e00032–13.

Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34(6):1037–1039.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid genome sequences. Genome Res. 27(5):757–767.

Werren JH. 1997. Biology of Wolbachia. Annu Rev Entomol. 42(1):587–609.

Werren JH, Baldo L, Clark ME. 2008. Wolbachia: master manipulators of invertebrate biology. Nat Rev Microbiol. 6(10):741–751.

Wu M, et al. 2004. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLos Biol. 2(3):e69–e341.

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(6):e38544.

Associate editor: Howard Ochman