A bioinformatics approach to identifying *Wolbachia* infections in arthropods

Jane Pascar Corresp., Christopher H. Chandler

1 Department of Biological Sciences, State University of New York at Oswego, Oswego, NY, United States

2 Department of Biology, Syracuse University, Syracuse, NY, United States

Corresponding Author: Jane Pascar
Email address: jpascar@oswego.edu

*Wolbachia* is the most widespread endosymbiont, infecting >20% of arthropod species, and capable of drastically manipulating the host’s reproductive mechanisms. Conventionally, diagnosis has relied on PCR amplification; however, PCR is not always a reliable diagnostic technique due to primer specificity, strain diversity, degree of infection and/or tissue sampled. Here, we look for evidence of *Wolbachia* infection across a wide array of arthropod species using a bioinformatic approach to detect the *Wolbachia* genes *ftsZ*, *wsp*, and the *groE* operon in next-generation sequencing samples available through the NCBI Sequence Read Archive. For samples showing signs of infection, we attempted to assemble entire *Wolbachia* genomes, and in order to better understand the relationships between hosts and symbionts, phylogenies were constructed using the assembled gene sequences. Out of the 34 species with positively identified infections, eight species of arthropod had not previously been recorded to harbor *Wolbachia* infection. All putative infections cluster with known representative strains belonging to supergroup A or B, which are known to only infect arthropods. This study presents an efficient bioinformatic approach for post-sequencing diagnosis and analysis of *Wolbachia* infection in arthropods.
A bioinformatics approach to identifying *Wolbachia* infection in arthropods

Jane A. Pascar¹,²*, Christopher H. Chandler¹

¹Department of Biological Sciences, State University of New York at Oswego, Oswego, NY, United States
²Present address: Department of Biology, Syracuse University, Syracuse, NY

* Corresponding Author: Jane Pascar (jpascar@oswego.edu)
Abstract

Wolbachia is the most widespread endosymbiont, infecting &gt;20% of arthropod species, and capable of drastically manipulating the host’s reproductive mechanisms. Conventionally, diagnosis has relied on PCR amplification; however, PCR is not always a reliable diagnostic technique due to primer specificity, strain diversity, degree of infection and/or tissue sampled. Here, we look for evidence of Wolbachia infection across a wide array of arthropod species using a bioinformatic approach to detect the Wolbachia genes ftsZ, wsp, and the groE operon in next-generation sequencing samples available through the NCBI Sequence Read Archive. For samples showing signs of infection, we attempted to assemble entire Wolbachia genomes, and in order to better understand the relationships between hosts and symbionts, phylogenies were constructed using the assembled gene sequences. Out of the 34 species with positively identified infections, eight species of arthropod had not previously been recorded to harbor Wolbachia infection. All putative infections cluster with known representative strains belonging to supergroup A or B, which are known to only infect arthropods. This study presents an efficient bioinformatic approach for post-sequencing diagnosis and analysis of Wolbachia infection in arthropods.

Introduction

Symbiotic relationships are ubiquitous in nature and can vary between parasitic, commensal, and mutualistic. Wolbachia is a diverse and widespread α-proteobacterium and obligatory endosymbiont (Werren, 1997a,b; Saridaki & Bourtzis, 2010). Wolbachia was first described in Culex pipiens (Hertig, 1936) and has since been identified in various clades of arthropods including Chelicerata (Werren & Windsor, 2000), Myriapoda (Mock et al., 2016), Crustacea (Bouchon, Rigaud & Juchault, 1998; Cordaux, Michel-Salzat & Bouchon, 2001; Cordaux et al., 2012), and Hexapoda (Werren & Windsor, 2000; Clark et al., 2001; Augustinos et al., 2011; Bing et al., 2014). Conservative estimates suggest that the frequency of Wolbachia infection in arthropods is at least 20% (Werren, 1997b; Werren & Windsor, 2000), while one study suggests a prevalence as high as 76% of arthropod species (Jeyaprakash & Hoy, 2000). Meta-analysis indicates that the infection distribution in the total number of species may be closer to 66% (Hilgenboecker et al., 2008).
Wolbachia is normally transmitted vertically, from mother to offspring, and can manipulate the host’s reproduction through five mechanisms: cytoplasmic incompatibility, parthenogenesis, male killing, feminization (Cordaux, Bouchon & Grève, 2011) and meiotic drive (Kageyama et al., 2017). However, there is evidence that Wolbachia can be horizontally transmitted (Vavre et al., 1999; Cordaux, Michel-Salzat & Bouchon, 2001; Raychoudhury et al., 2009; Kraaijeveld et al., 2011). Recently, discrete reciprocal benefits provided by infection have been observed including a positive impact on host immunity (Teixeira, Ferreira & Ashburner, 2008; Osborne et al., 2009), immunocompetence (Braquart-Varnier et al., 2008), fecundity (Weeks et al., 2007), and metabolic activity (Darby et al., 2012).

Currently, all Wolbachia strains are classified as a single species, with further classification into at least sixteen supergroups, A-Q (Lo et al., 2007; Lindsey et al., 2016a). The four most well studied clades are supergroups A-D. Supergroups A and B are monophyletic and are the most common supergroups known to infect arthropods, while supergroups C and D infect filarial nematodes (Gerth et al., 2014). Supergroup G was discovered to be a recombinant between supergroups A and B; thus it is no longer considered a distinct lineage (Baldo & Werren, 2007). Supergroups E-Q infect a variety of hosts including nematodes, springtails, termites, fleas, aphids, and mites (Lo et al., 2002; Casiraghi et al., 2005; Ros et al., 2009; Haegeman et al., 2009; Augustinos et al., 2011; Bing et al., 2014; Glowska et al., 2015).

Wolbachia has a relatively small genome at about 0.9-1.5 Mbp. Historically, Wolbachia infection was diagnosed using 16S rRNA sequences; however, strains range in divergence from 0.2% to 2.6%, and when used independently, 16S provides limited information for inferring phylogenetic relationships (O’Neill et al., 1992). wsp, ftsZ and the groE operon are all protein-encoding genes used for the detection and phylogenetic analysis of Wolbachia (Van Borm et al., 2003). The ftsZ gene is involved in cell division and is highly conserved in unculturable bacteria species (Holden, Brookfield & Jones, 1993), but regions that are relatively higher in divergence make it a candidate for better phylogenetic resolution allowing the distinction between supergroups A and B to become apparent (Werren, 1997a). The wsp gene, which codes for the surface protein WSP in Wolbachia, shows an even higher variability and faster evolutionary rate than 16S or ftsZ and can be used in identifying groups and strains of Wolbachia (Zhou, Rousset & O’Neil, 1998; Braig et al., 1998), but also displays recombination, which can be misleading when used in phylogenetic analyses (Baldo & Werren, 2007). The groE-homologous operon has
been noted as another candidate for resolving strain taxonomy (Masui, Sasaki & Ishikawa, 1997). Only a single copy of the operon exists in the genome and it includes the genes that encode the heat shock proteins GroES and GroEL, which are separated by a non-coding intergenic region that is thought to be faster evolving than either of the coding regions (Masui, Sasaki & Ishikawa, 1997).

With the use of antibiotics, *Wolbachia* infections in some species have been cured and the phenotypic changes that are induced by infection are consequently reversed (Stouthamer, Luck & Hamilton, 1990; Bourtzis et al., 1994; Giordano, Jackson & Robertson, 1997). More recently, *Wolbachia* has been proposed as a natural solution to controlling the spread of vector-borne diseases like malaria, yellow fever, and dengue (Hoffmann et al., 2011; Walker & Moreira, 2011; Baldini et al., 2014b). Arthropods are present in nearly every habitat on Earth and they play important ecological roles in a variety of niches. With an estimated 2.4-10.2 million species of arthropods (Ødegaard, 2000) it is important to quantify the prevalence and distribution of *Wolbachia* infection.

*Wolbachia* infections are typically diagnosed via polymerase chain reaction (PCR), using *Wolbachia*-specific primers. However, PCR-based tests may produce false positives or false negatives, depending on the strain of *Wolbachia* and the presence of other related bacterial symbionts (Simões et al., 2011). A metagenomics-based approach can also be useful for characterizing microbiomes, including looking for *Wolbachia* and other symbionts (e.g., (Dittmer & Bouchon, 2018), and can even provide whole-genome sequence information for the symbiont (e.g., (Salzberg et al., 2005; Richardson et al., 2012; Saha et al., 2012; Campana, Robles García & Tuross, 2015; Derks et al., 2015; Wang & Chandler, 2016; Lindsey et al., 2016b; Gerth & Hurst, 2017). While performing a high-throughput sequencing-based screen for *Wolbachia* involving hundreds of different species would require a huge sampling effort and could be cost-prohibitive, screening existing sequence datasets generated for other projects offers a powerful opportunity to diagnose novel infections and better characterize variation in symbionts.

Here, using publicly accessible next-generation sequencing data available in the NCBI Sequence Read Archive (SRA), we looked for evidence of *Wolbachia* infection in a diverse assemblage of arthropod species. We present methods for bioinformatically identifying *Wolbachia* infections in genomic samples. We then used these sequence data to assemble a draft
genome sequence for each *Wolbachia* isolate and reconstruct the phylogenetic relationships among the identified *Wolbachia* strains. Using this approach, we uncover novel *Wolbachia* infections, as well as find possible evidence for horizontal transfer of *Wolbachia* between hosts and parasites. These results illustrate how existing genetic databases can provide a wealth of information on symbiotic microbes as a byproduct of host sequencing.

## Materials & Methods

### Retrieving Data

All samples tested are available through the NCBI Sequence Read Archive (SRA) (Supplementary Table S1). To identify samples for testing, all accession numbers that matched the criteria of Arthropoda genomic DNA were sent to the NCBI Run Selector (as of January 2017). In the Run Selector samples were selected based on the criteria that they were run on an Illumina platform, have a genomic library source, a random library selection and the library layout is paired. Transcriptome samples were excluded because of the possibility that some RNA preparation methods may select against bacterial RNA (e.g., poly-A enrichment (Westermann, Gorski & Vogel, 2012)) thus increasing the likelihood of false negatives and because assembling *Wolbachia* genomes would be impossible with these data. Similarly, targeted sequencing (e.g., RAD-seq) samples were excluded due to the possibility that the sequences used for detecting *Wolbachia* infections might be excluded during the library preparation process. Only paired-end datasets were considered in order to facilitate whole-genome assembly for positive samples, but there were relatively few species (only 22) in the database with single-end datasets that otherwise met our criteria. Every species that had a sample that met our criteria was chosen for sampling. Some species were over-represented in the number of runs that are available in the SRA; depending on the number of samples available in the SRA, an appropriate amount to include in our dataset was determined on a case-by-case basis (Supplementary Table S2). Fastq-dump v. 2.8.0 from the SRA Toolkit (NCBI SRA) was used to download, at most, 5x10^7 reads from each accession.

### Diagnosing *Wolbachia* Infection

Magic-Blast v1.1.0 (NCBI) was used to compare the SRA reads to selected reference *wsp*, *ftsZ*, and *groE* operon sequences isolated from *Wolbachia* samples that are representative of
supergroups A-D (Table 1). A custom R script identified SRA samples where there were matches at least 98 bp in length, ≥95% identity to one or more of the reference genes, and with three or more matching sequence reads. All samples that met these criteria were called *Wolbachia* positive samples.

To look for previous reports of *Wolbachia* infection in the species that tested positive, first Google Scholar was used. [species] + *Wolbachia* was used for the search terms. If no published results were found, next we used NCBI GenBank with the same search parameters to look for deposited sequences that may be unpublished that would indicate that *Wolbachia* had been found in the host species previously.

**Assembling the *Wolbachia* Gene and Genome Sequences**

From all the samples that tested positive (Supplementary Table S3) if there were more than 3 samples from one species a maximum of 3 samples were chosen for downstream analysis (Table S4). Velvet v1.2.10 (Zerbino & Birney, 2008) was used to separately assemble the *wsp*, *ftsZ*, and *groE* sequences for each biological sample using the sequence reads that aligned to each gene in the previous step. It was run for kmer values of 21, 31, 41, and 51, using the automatic coverage cutoff flag. To select the optimal assembly of each gene for each sample, we performed BLASTn v2.28 (Altschul et al., 1990), which searched against a database made of each respective reference gene (Table 1). BEDTools v2.25.0 (Quinlan & Hall, 2010) and a custom script was used to parse the BLASTn results for the single longest contig matching each gene from each sample.

To assemble draft genomes for each *Wolbachia* isolate we identified, an iterative bait-and-assemble approach was used. Independent SRA experiments or runs from the same BioSample were first combined into a single dataset. For each sample, the mirabait tool from MIRA v4.0.2 (Chevreux, Wetter & Suhai, 1999) was then used to extract all reads from the full dataset that shared at least one kmer with at least one of seven reference *Wolbachia* genomes representing *Wolbachia* isolates from insects and nematodes (wPip, GCF_000073005.1; wMel, GCF_000008025.1; wNo, GCF_000376585.1; wRi, GCA_000022285.1; wVol, GCF_000530755.1; wCle, GCF_000829315.1; wTpre, GCF_001439985.1), using $k=31$. These reads, and their corresponding paired-end partners, were assembled using SPAdes 3.11.1 (Bankevich et al., 2012). All resulting contigs were then aligned to the reference *Wolbachia* genomes using dc-megablast 2.7.0+ (Camacho et al., 2009), and any contig that matched any of
the reference genomes with an e-value of $10^{-10}$ or better, alignment length of at least 100bp, and percent identity of at least 70%, was retained. This process was then repeated for a total of five iterations, in each cycle using mirabait to identify reads sharing one or more kmers with the last set of assembled contigs, re-assembling these putatively *Wolbachia*-derived reads, and retaining any of the newly assembled contigs that show similarity to a *Wolbachia* reference genome in BLAST searches. The quality of each final assembly was evaluated using QUAST v4.4 (Gurevich et al., 2013) and BUSCO v3.0.2b (Simão et al., 2015) with the Bacteria odb9 reference gene set. Finally, we mapped all sequencing reads from each associated BioSample (not just those used for the assembly process) to the corresponding assembly using bwa mem v.0.7.17 (Li, 2013), and then used the sambamba depth command (Tarasov et al., 2015) to extract coverage information for each assembled contig over 400 bp in length, excluding 150bp from the ends of the contigs (where coverage tends to drop off because reads extending beyond the contig may fail to map successfully).

**Phylogenetic Analysis**

We first constructed phylogenies using the assembled *ftsZ* and *groE* sequences, as well as from a concatenated dataset of both genes; *wsp* was excluded from phylogenetic analyses because of its high frequency of recombination (Baldo & Werren, 2007). *Wolbachia* gene sequences representing *ftsZ* and the *groE* operon from other studies where the supergroup classification was determined were used as control samples; in this analysis, we included only reference sequences where both genes had been sequenced from the same biological sample (Table 2). MAFFT v7.310 (Katoh & Standley, 2013) was used to align the sequences for each respective gene. Samples that lacked sufficient length of matching base pairs (at least 800 bp in total across both genes) were discarded from downstream analysis. GBlocks v0.91b (Castresana, 2000) removed the poorly aligned portions of the sequences from each gene alignment using the default parameters. MEGA v7.0 (Kumar, Stecher & Tamura, 2016) was used to construct phylogenies using maximum likelihood. The model for which the phylogenies were constructed was chosen according to MEGA's suggestion for best fit based on the lowest Bayesian information criterion (BIC) (Table S5). Node support was assessed by bootstrapping with 1,000 replicates.

We also constructed phylogenies based on whole-genome data from a subset of the assemblies which appeared the most complete based on the BUSCO assessment. For these
phylogenies, we used REALPHY (Bertels et al., 2014), to align genome sequences and identify loci for inclusion in the phylogenetic analysis, using as a reference the seven Wolbachia genomes used in the assembly process and merging the reference alignments with the default parameters. We then performed phylogenetic analysis by maximum likelihood in RAxML v8.2.11 (Stamatakis, 2014) using the TVM+I+G model as selected by ModelTest-NG v0.1.2 (the successor to jModelTest); (Darriba et al., 2012) using AIC. The RAxML analysis included 100 independent replicate searches for the best-scoring tree and 200 bootstrap replicates to assess node support.

Results

Diagnosing Wolbachia infections from publicly available sequence data

2,545 individual 'runs' from the SRA, representing 288 species and subspecies were tested for Wolbachia (Supplementary Table S1). Of those, 173 runs from 34 unique species tested positive for the selected reference Wolbachia genes (Supplementary Table S3). That is, 11.8% of species tested positive for Wolbachia in at least one sample and only 6.8% of all SRA runs tested positive. All samples that tested positive were from samples that are in the class Insecta and representative of five orders: Coleoptera, Diptera, Hymenoptera, Hemiptera, and Lepidoptera. According to our literature search eight of these species have not previously been confirmed to have Wolbachia infections—Bembidion lapponicum, Ceratina calcarata, Delias oriaia, Diachasma alloaeum, Diploeciton nevermanni, Ecitophyla simulans, Gerris buenoi and Isocolus centaureae (Table 3).

Assembling Wolbachia genomes

In total, we assembled draft genomes for 51 Wolbachia isolates (Table 4), including at least one for each of the 34 unique host species. There were only two cases in which the assembly was substantially smaller than the expected genome size. In one of those (Biorhiza pallida 3), infection was confirmed in independent biological samples, and in the other (Mycopsylla proxima) the small assembly probably resulted from the small size of the input dataset. The rest of these assemblies appeared nearly complete, with total assembly sizes of at least 1 Mb and high numbers of BUSCO reference genes represented by a single gene in the assembly. All assemblies were missing at least 13 of the BUSCO reference genes.
We also sought to determine whether each sample that tested positive was likely to represent an actual *Wolbachia* infection, or the result of *Wolbachia* sequences horizontally transferred into the host genome. If the sequencing depth of the *Wolbachia*-like contigs in the assembly differs substantially from the sequencing depth of the host genome, then horizontal transfer can be ruled out. However, performing whole-genome assembly with every sequence dataset to estimate the sequencing depth of the host genome was computationally time-consuming, and our attempts to estimate sequencing depth more rapidly by counting \( k \)-mers in the raw data were unsuccessful in most cases because of low sequencing depth. Therefore, we obtained estimates of the genome size of host species from other sources, such as draft assemblies available at NCBI (Table 4), when available; although draft assemblies can differ substantially in size from actual genome sizes, for our purposes this should be a reasonable approximation. We then estimated the expected sequencing depth of the host by dividing the total amount of sequencing data by the estimated host genome size. Although genome size data on some host species was lacking, large differences in sequencing depth between the host and endosymbiont support active *Wolbachia* infections in several species, including *Anopheles gambiae*, *Diabrotica virgifera*, *Diachasma alloeum*, *Diaphorina citri*, several *Drosophila* species, *Homalodisca vitripennis*, *Rhagoletis pomonella* and *R. zephyria*, *Trichogramma pretiosum*, *Callosobruchus chinensis*, *Ceratina calcarata*, and *Dactylopius coccus* (Table 4).

In a few cases, there was evidence of multiple infections in a single sample. This evidence included an unusual number of duplicated BUSCO reference genes in the assembly (e.g., *Homalodisca vitripennis* 1), the presence of multiple peaks in the coverage distribution histogram (e.g., *Callosobruchus chinensis*), assembly sizes much larger than previously sequenced Wolbachia genomes (e.g., *Dactylopius coccus*), or some combination of these (Table 4).

**Wolbachia Phylogeny**

All phylogenetic trees based on individual or concatenated datasets using the *ftsZ* and *groE* sequences show two distinct branches representing supergroups A and B (Figure 1; Supplementary Figures S1-S2). The tree resulting from the concatenated dataset has the most robust bootstrap support for most clades. Positive control samples that were included in the phylogeny cluster with other control samples of the same known supergroup. Of the species where *Wolbachia* had been previously unidentified, according to all trees, the strains isolated
from *D. allosum* falls within supergroup A, while the *B. lapponicum, I. centaureae, G. buenoi, D. nevermanni* and *D. oraia* isolates all fall within supergroup B (Figures 1-2, Supplementary Figures S1-S2).

The phylogeny generated from whole-genome sequencing data (Figure 2) was similar in overall topology to the trees based on *ftsZ* and *groE*, with two clear clades representing supergroups A and B, but with higher bootstrap support for most branches.

**Discussion**

**Observed Low Infection Rates**

While *Wolbachia* is estimated to infect between 20-76% of arthropod species (Werren, Windsor & Guo, 1995; Jeyaprakash & Hoy, 2000), in this set of data only 11.8% of species tested positive. Given the source, this low rate of infection can be hypothesized to be the result of five possible scenarios: (1) Underrepresentation in the amount of data available per host species. For example, only 43 out of the 288 (14.9%) species and subspecies tested had ≥10 samples available in the SRA that met the criteria of this study (Supplementary Table S2). When >100 individuals are tested for *Wolbachia*, results are skewed towards finding a positive sample (Hilgenboecker et al., 2008). (2) Bias in the source of the samples. Sources vary between wild-caught individuals, lab stocks, and unreported sources. Since the phenotypic consequences of *Wolbachia* are well established, if uninfected individuals are needed for a study they may be selectively chosen (see Đorđević et al., 2017; Becking et al., 2017), or the researchers may even actively treat infections with antibiotics (Dobson & Rattanadechakul, 2001; Casiraghi et al., 2002; Koukou et al., 2006) or increased rearing temperature. In those cases the sequencing data will consequently test negative for *Wolbachia* using the methods employed here. (3) Tissue sampled. In some species infection has only been detectable in the gonads, indicating that infection density in somatic tissue may be variable or low (Dobson et al., 1999). For many samples in the SRA, specific tissue has not been indicated. (4) Bioinformatic removal of bacterial contaminants. Even if *Wolbachia* is sequenced with the host’s DNA, the researcher may have eliminated these reads bioinformatically before depositing the reads as relatively standard practice in sequence processing (Kunin et al., 2008; Schmieder & Edwards, 2011; Derks et al., 2015). (5) False negatives. It is possible that some infections may have been missed
due to the limited set of available reference genes; more divergent strains might not have been detected in these analyses.

**Strain Supergroup Affiliation**

For 13 of the species that tested positive, previous information was available about supergroup affiliations of *Wolbachia* strains that have been found to infect them (Table 3). Our results are mostly consistent with previously reported phylogenetic relationships. Previously, *C. chinensis, D. coccus, and D. simulans* have been found to be infected with A and/or B strains (Kondo, Shimada & Fukatsu, 1999; Kondo et al., 2002; Riegler et al., 2004; Ellegaard et al., 2013; Ramirez-Puebla et al., 2016). Here, evidence of both A and B supergroup strains was found in *D. simulans* (Figure 1), though the whole-genome phylogeny was somewhat inconsistent here, suggesting possible recombination for some genes. Moreover, while the single-gene phylogenies suggested that the endosymbionts of these *C. chinensis* and *D. coccus* samples were members of supergroups A and B, respectively, the whole-genome assemblies for both of these endosymbionts contained strong evidence of dual infections, so we cannot rule out the presence of both A and B supergroup strains in these samples.

*Wolbachia* infection has also been documented prior to this study in *B. pallida, P. aegeria, and P. sp. PSW-54* but the supergroup relationships were not reported (Subandiyah et al., 2000; Rokas et al., 2001; Russell et al., 2012; Kautz, Rubin & Moreau, 2013; Baldini et al., 2014b). Our concatenated results suggest that a supergroup A strain infects *B. pallida* and *P. sp. PSW-54*, while a B strain infects *P. aegeria*.

Two species showed a different supergroup strain than what has been previously reported—*D. spinosa* and *A. gambiae*. *D. spinosa* has previously been identified to harbor a supergroup A strain, but here we discovered an infection that clusters within supergroup B. It may be possible for *D. spinosa* to harbor both A and B strains since other species in the genus have been shown to have supergroup B infections (Plantard et al., 1999).

Particularly notable is our identification of a supergroup B strain in *A. gambiae*. Anopheline mosquitoes were once thought to lack infection by *Wolbachia* in nature (Kittayapong et al., 2000; Ricci et al., 2002; Rasgon & Scott, 2004), though they are capable of experimental infection in the lab (Hughes et al., 2011). However, there have been recent reports of natural infections in wild populations (Baldini et al., 2014a; Gomes et al., 2017). In particular, a supergroup A strain was found to infect *A. gambiae* mosquitoes in Mali that reduces the
transmission of the malaria parasite (Gomes et al., 2017). The strain identified here clearly
belongs to supergroup B, and is related to strains infecting Hymenoptera and Lepidoptera, rather
than fleas (Siphonaptera) like the previously identified supergroup A strain. In addition, other
recent surveys have found evidence of diverse *Wolbachia* strains, including supergroup B strains,
within the *A. gambiae* species complex (Ayala et al., 2018; Jeffries et al., 2018). Combined,
these results suggest that the diversity of *Wolbachia* infections in *Anopheles* may be currently
underappreciated. Importantly, we have good evidence that the sample here is an actual
*Wolbachia* infection rather than an integrated piece of *Wolbachia* DNA in the host genome. First,
we assembled a nearly complete *Wolbachia* genome from this dataset; more importantly, given
that the *A. gambiae* genome is roughly 280 Mb (Holt, 2002), and this dataset contained roughly
9.1 Gb of raw sequence reads, we would expect to have roughly ~32x coverage of the host
genome, but the *Wolbachia* genome had only ~9x coverage, suggesting that *Wolbachia* DNA
was present at lower densities in this sample than the host DNA.

The assembled *ftsZ* and *groE* sequences from *C. calcarata, E. simulans, M. fici,* and *M. proxima* assemblies were too short to be included in our individual gene-based phylogenetic
reconstruction; the infection density, and thus the sequencing coverage, for these species may
have been too low to yield reliable assemblies for these genes. We were able to assign the *E.
simulans* infection to supergroup B based on its draft genome sequence, but the supergroup
relationships for the others are still unknown. Previously, *Wolbachia* sequence information has
been isolated in *M. fici,* and *M.proxima* (Fromont et al., unpublished data; Table 3) but
supergroup affiliation was not suggested. According to our literature search this is the first
detection of *Wolbachia* in *C. calcarata* and *E. simulans.* The other six previously unidentified
species infections were included in the phylogeny. *D. alloeum* clustered with known supergroup
*A* infections while *B. lapponicum, D. oraia, D. nevermanni, G. buenoii, I. centaureae* isolate
clustered within the supergroup B clade.

Finally, our phylogeny also offers some hints into possible mechanisms of horizontal
transmission of *Wolbachia* infections. In particular, the strain identified here infecting
*Diachasma alloeum* is closely related to the strain found in *Rhagoletis pomonella* (as well as *D.
melanogaster* and *D. yakuba*). This is intriguing because *D. alloeum* is a parasitoid wasp that
uses *R. pomonella* and *R. mendax* as its host (Maier, 1981; Stelinski, Pelz & Liburd, 2004),
suggesting that this may represent a natural horizontal transfer of *Wolbachia* from one lineage to
another; previous studies have found evidence of horizontal transmission between predators and
prey or hosts and parasites (Heath et al., 1999; Le Clec'h et al., 2013). However, contamination
by host material in parasitoid samples, or vice versa, could also explain this outcome, so this
result should be interpreted cautiously until this path of transmission can be experimentally
confirmed.

**Multiple Infections and Integration of Wolbachia into the Host Genome**

Double (Perrot-Minnot, Guo & Werren, 1996; Narita, Nomura & Kageyama, 2007) and
even triple *Wolbachia* infections (Rousset, Braig & O’Neill, 1999; Kondo et al., 2002) have been
reported in arthropod populations and individuals, both naturally and through experimental
injection. The initial screening methods presented here are not capable of identifying multiple
infections because we only looked for a positive or negative test result and then used only the
single longest contig for phylogenetic construction. In conventional PCR there is a tradeoff
between specificity and sensitivity of primers; additionally no one primer is capable of
identifying *Wolbachia* in all samples (Simões et al., 2011). PCR is useful in initial infection
confirmation but sequencing is usually necessary to confirm group relationships. Techniques
used to identify multiple infections currently include quantitative PCR with highly specific
primers (Kondo et al., 2002; Narita, Nomura & Kageyama, 2007), cloning and sequencing
(Jamnongluk et al., 2002), and Southern hybridization (Perrot-Minnot, Guo & Werren, 1996).

We were able to identify evidence of possible multiple infections through genome
assembly. In some cases, the assembly was approximately double the expected size, contained a
large number of duplicated genes, or showed evidence of multiple peaks in a coverage
histogram, all of which are signs of infection by multiple, independent strains. Again, these
results should be interpreted cautiously pending experimental validation. For instance, some of
the multiply infected samples consisted of pooled DNA from multiple individuals (e.g.,
*Drosophila yakuba* and *Diabrotica virgifera*), so the "multiple" infection might simply result
from different individuals in the sample harboring different endosymbiont strains. Nevertheless,
these results show that high-throughput sequencing can be a powerful way to detect multiple
infections, especially when a priori sequence information for designing strain-specific primers is
unavailable.

A related issue is that *Wolbachia* DNA is frequently integrated into host genomes (Vavre
et al., 1999; Leclercq et al., 2016); in some cases, these insertions even consist of nearly whole
Wolbachia genome sequences (Dunning Hotopp et al., 2007). This complicates our analyses because some of the identified "infections" could actually be Wolbachia DNA integrated into the host genome; in fact, horizontally transferred Wolbachia DNA has already been identified in four orders which are all represented by the positive results in this study, Coleoptera, Diptera, Hemiptera, and Hymenoptera (Dunning Hotopp et al., 2011). We were able to rule out horizontally transferred DNA in some, but not all, cases of positive samples, using sequencing depth information; if the sequencing depth of the assembled Wolbachia contigs differs from the sequencing depth of the host's nuclear DNA, that suggests a true, active infection. True infections could also be validated experimentally when necessary, for example, using fluorescence in situ hybridization (Hughes et al., 2011). Either way, horizontally transferred Wolbachia DNA would still indicate that a species at least had a history of infection at some point in the past.

This work shows that it is often possible to assemble draft genomes of endosymbionts from host DNA, similar to previous studies in which Wolbachia genomes were assembled from sequencing host organisms (Ghedin et al., 2004; Salzberg et al., 2005; Richardson et al., 2012; Saha et al., 2012; Campana, Robles Garcia & Tuross, 2015; Derks et al., 2015; Lindsey et al., 2016b), even when the endosymbiont was not the focus or original reason for performing the sequencing in the first place. Although they may be fragmented, these draft genomes can still provide valuable information about the phylogenetics and evolution of the endosymbiont. While Wolbachia is relatively well studied, there are many other endosymbionts that have received less attention, such as some Spiroplasma, Cardinium, Arsenophonus, and Flavobacetrion species (Duron et al., 2008), and others await discovery. This study shows that extensive field sampling may not even be necessary to get a better understanding of the diversity of these endosymbionts; the sequencing data are probably already available in public databases. With the right reference databases and metagenomics software, there is a lot of potential to learn more about these endosymbionts just from already existing resources.

Conclusions and Recommendations

Wolbachia is a well-known endosymbiont of many arthropod species and while standard Wolbachia diagnostic techniques utilize various Wolbachia primers to confirm infection via PCR (Simões et al., 2011) there are trade-offs that limit large scale surveys. Here, we present a
method to identify *Wolbachia* bioinformatically using publicly accessible host raw sequencing data. In eight arthropod species, *Wolbachia* was identified where infection has not previously been reported, and in 27 other arthropod species infection was confirmed. Isolates of *Wolbachia* from positive samples all clustered within either supergroups A or B, and for seven of the newly identified hosts we identified the supergroup of the strain. From these isolates we assembled draft *Wolbachia* genomes, which provided robustly supported phylogenetic information as well as information about potential HGT events or signs of multiple infection.

These results highlight the importance of depositing raw sequencing datasets to public archives like the NCBI SRA and the value that they have in studying endosymbionts. At the same time, we offer some suggestions for best practices when depositing sequence data into public archives to maximize its usefulness for future researchers (Wilkinson, Dumontier & and others, 2016; Griffin et al., 2017). First, we encourage everyone performing high-throughput sequencing to deposit their data into public databases like the NCBI SRA, where it can easily be searched and accessed, as opposed to depositing only in smaller, taxon-specific databases or personal/lab web sites. Second, data should be minimally filtered; while "contaminant" sequences like endosymbiont DNA may be a nuisance to those who generated the data, they may be of interest to others. Finally, all sequence data should be accompanied by as much metadata as possible. Without this information, interpreting results can be difficult. For example, many of the sequences we used in this study lacked detailed information about the source of the DNA in the associated BioSample entries (e.g., whether it came from a lab strain or wild-caught specimens, its geographic origin if field collected, whether it was from a single individual or a pooled sample, whether the specimen was male or female, whether it was a whole body or specific tissues, etc.). Including this information would have helped us better understand possible biases in the dataset, such as how well the results may reflect the frequency of infection in natural populations, or whether a sample might give a false negative result because *Wolbachia* is not present at high densities in the tissues sampled for DNA.

**Acknowledgments**

We thank P Brannock, S Borrelli, and the editor and three anonymous reviewers for their helpful feedback and suggestions on earlier versions this manuscript. We also thank V Buonaccorsi, C
Walls, and GCAT-SEEKquence for computing support, as well as the National Center for Genome Analysis Support at Indiana University.

Additional citations from Tables and Figure Legends to incorporate into References section below:

(Ros et al., 2012) (Rowley, Raven & McGraw, 2004) (Baldo et al., 2006) (Wiwatanaratnabutr et al., 2009) (Tsai et al., 2004) (Mateos et al., 2006) (Werren & Bartos, 2001) (Haine, Pickup & Cook, 2005) (Kyei-Poku et al., 2006) (Pintureau et al., 2000) (Schilthuizen & Stouthamer, 1998) (Zha et al., 2014) (Hurst et al., 1999) (Schilthuizen, Honda & Stouthamer, 1998) (Jiggins et al., 2002) (Werren, Zhang & Guo, 1995) (Bazzocchi et al., 2000) (Verne et al., 2007) (Paraskevopoulos et al., 2006) (Cordaux et al., 2008) (Schuler et al., 2011) (Rogers & Backus, 2014) (Husnik & McCutcheon, 2016) (Kikuchi & Fukatsu, 2003) (Frost et al., 2010) (Abe & Miura, 2002) (Rokas et al., 2002) (Kodandaramaiah et al., 2011)(Lindsey et al., 2016b)

References

Abe Y., Miura K. 2002. Does Wolbachia Induce Unisexuality in Oak Gall Wasps? (Hymenoptera: Cynipidae). *Annals of the Entomological Society of America* 95:583–586. DOI: 10.1603/0013-8746(2002)095[0583:DWIUIO]2.0.CO;2.

Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Augustinos AA., Santos-Garcia D., Dionyssopoulou E., Moreira M., Papapanagiotou A., Scarvelakis M., Doudoumis V., Ramos S., Aguiar AF., Borges PA V., Khadem M., Latorre A., Tsiamis G., Bourtzis K. 2011. Detection and Characterization of Wolbachia Infections in Natural Populations of Aphids: Is the Hidden Diversity Fully Unraveled? *PLoS ONE* 6:e28695. DOI: 10.1371/journal.pone.0028695.

Ayala D., Akone-Ella O., Rahola N., Kengne P., Ngangue MF., Mezeme F., Makanga BK., Costantini C., Simard F., Prugnolle F., Roche B., Duron O., Paupy C. 2018. Natural Wolbachia infections are common in the major malaria vectors in Central Africa. *bioRxiv*. DOI: 10.1101/343715.

Baldini F., Segata N., Pompon J., Marcenac P., Robert Shaw W., Dabiré RK., Diabaté A.,
Levashina EA., Catteruccia F. 2014a. Evidence of natural Wolbachia infections in field populations of Anopheles gambiae. Nature Communications 5. DOI: 10.1038/ncomms4985.

Baldini F., Segata N., Pompon J., Marценac P., Shaw WR., Dabiré RK., Diabaté A., Levashina EA., Catteruccia F. 2014b. Evidence of natural Wolbachia infections in field populations of Anopheles gambiae. Nature Communications 5. DOI: 10.1038/ncomms4985.

Baldo L., Hotopp JCD., Jolley KA., Bordenstein SR., Biber SA., Choudhury RR., Hayashi C., Maiden MCJ., Tettelin H., Werren JH. 2006. Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Applied and Environmental Microbiology 72:7098–7110. DOI: 10.1128/AEM.00731-06.

Baldo L., Werren JH. 2007. Revisiting Wolbachia Supergroup Typing Based on WSP: Spurious Lineages and Discordance with MLST. Current Microbiology 55:81–87. DOI: 10.1007/s00284-007-0055-8.

Bankevich A., Nurk S., Antipov D., Gurevich AA., Dvorkin M., Kulikov AS., Lesin VM., Nikolenko SI., Pham S., Prjibelski AD., Pyshkin A V., Sirotkin A V., Vyahhi N., Tesler G., Alekseyev MA., Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19:455–477. DOI: 10.1089/cmb.2012.0021.

Bazzocchi C., Jamnongluk W., O’Neill SL., Anderson TJC., Genchi C., Bandi C. 2000. wsp Gene Sequences from the Wolbachia of Filarial Nematodes. Current Microbiology 41:96–100. DOI: 10.1007/s002840010100.

Becking T., Giraud I., Raimond M., Moumen B., Chandler C., Cordaux R., Gilbert C. 2017. Diversity and evolution of sex determination systems in terrestrial isopods. Scientific Reports 7:1084. DOI: 10.1038/s41598-017-01195-4.

Bertels F., Silander OK., Pachkov M., Rainey PB., Van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. Molecular Biology and Evolution 31:1077–1088. DOI: 10.1093/molbev/msu088.

Bing X-L., Xia W-Q., Gui J-D., Yan G-H., Wang X-W., Liu S-S. 2014. Diversity and evolution of the Wolbachia endosymbionts of Bemisia (Hemiptera: Aleyrodidae) whiteflies. Ecology and Evolution 4:2714–2737. DOI: 10.1002/ece3.1126.

Van Borm S., Wenseleers T., Billen J., Boomsma J. 2003. Cloning and sequencing of wsp encoding gene fragments reveals a diversity of co-infecting Wolbachia strains in
Acromyrmex leafcutter ants. *Molecular Phylogenetics and Evolution* 26:102–109. DOI: 10.1016/S1055-7903(02)00298-1.

Bouchon D., Rigaud T., Juchault P. 1998. Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society B: Biological Sciences* 265:1081–1090. DOI: 10.1098/rspb.1998.0402.

Bourtzis K., Nirgianaki A., Onyango P., Savakis C. 1994. A prokaryotic dnaA sequence in Drosophila melanogasten Wolbachia infection and cytoplasmic incompatibility among laboratory strains. *Insect Molecular Biology* 3:131–142. DOI: 10.1111/j.1365-2583.1994.tb00160.x.

Braig HR., Zhou W., Dobson SL., O’Neill SL. 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont Wolbachia pipiensis. *Journal of bacteriology* 180:2373–8.

Braquart-Varnier C., Lachat M., Herbinière J., Johnson M., Caubet Y., Bouchon D., Sicard M. 2008. Wolbachia Mediate Variation of Host Immunocompetence. *PLoS ONE* 3:e3286. DOI: 10.1371/journal.pone.0003286.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:1. DOI: 421

Campana MG., Robles García NM., Tuross N. 2015. America’s red gold: Multiple lineages of cultivated cochineal in Mexico. *Ecology and Evolution* 5:607–617. DOI: 10.1002/ece3.1398.

Casiraghi M., Bordenstein SR., Baldo L., Lo N., Beninati T., Wernegreen JJ., Werren JH., Bandi C. 2005. Phylogeny of Wolbachia pipiensis based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the Wolbachia tree. *Microbiology* 151:4015–4022. DOI: 10.1099/mic.0.28313-0.

Casiraghi M., McCall J., Simoncini L., Kramer L., Sacchi L., Genchi C., Werren J., Bandi C. 2002. Tetracycline treatment and sex-ratio distortion: a role for Wolbachia in the moultng of filarial nematodes? *International Journal for Parasitology* 32:1457–1468. DOI: 10.1016/S0020-7519(02)00158-3.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in
phylogenetic analysis. *Molecular Biology and Evolution* 17:540–552. DOI:
10.1093/oxfordjournals.molbev.a026334.

Chevreux B., Wetter T., Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)* 99:45–56.

Clark TL., Meinke LJ., Skoda SR., Foster JE. 2001. Occurrence of Wolbachia in Selected Diabroticite (Coleoptera: Chrysomelidae) Beetles. *Annals of the Entomological Society of America* 94:877–885.

Le Clec’h W., Chevalier FD., Genty L., Bertaux J., Bouchon D., Sicard M. 2013. Cannibalism and Predation as Paths for Horizontal Passage of Wolbachia between Terrestrial Isopods. *PLoS ONE* 8. DOI: 10.1371/journal.pone.0060232.

Cordaux R., Bouchon D., Grève P. 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends in Genetics* 27:332–341. DOI: 10.1016/j.tig.2011.05.002.

Cordaux R., Michel-Salzat A., Bouchon D. 2001. Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. *Journal of Evolutionary Biology* 14:237–243. DOI: 10.1046/j.1420-9101.2001.00279.x.

Cordaux R., Pichon S., Ben Afia Hatira H., Doublet V., Grève P., Marcadé I., Braquart-Varnier C., Souty-Grosset C., Charfi-Cheikhrouha F., Bouchon D. 2012. Widespread Wolbachia infection in terrestrial isopods and other crustaceans. *ZooKeys* 176:123–131. DOI: 10.3897/zookeys.176.2284.

Cordaux R., Pichon S., Ling A., Pérez P., Delaunay C., Vavre F., Bouchon D., Grève P. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Molecular Biology and Evolution* 25:1889–1896. DOI: 10.1093/molbev/msn134.

Darby AC., Armstrong SD., Bah GS., Kaur G., Hughes MA., Kay SM., Koldkjær P., Rainbow L., Radford AD., Blaxter ML., Tanya VN., Trees AJ., Cordaux R., Wastling JM., Makepeace BL. 2012. Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Research* 22:2467–2477. DOI: 10.1101/gr.138420.112.

Darriba D., Taboada GL., Doallo R., Posada D. 2012. JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9:772. DOI: 10.1038/nmeth.2109.
Derks MFL., Smit S., Salis L., Schijlen E., Bossers A., Mateman C., Pijl AS., de Ridder D., Groenen MAM., Visser ME., Megens H-J. 2015. The Genome of Winter Moth (Operophtera brumata) Provides a Genomic Perspective on Sexual Dimorphism and Phenology. Genome Biology and Evolution 7:2321–2332. DOI: 10.1093/gbe/evv145.

Dittmer J., Bouchon D. 2018. Feminizing Wolbachia influence microbiota composition in the terrestrial isopod Armadillidium vulgare. Scientific Reports 8:6998. DOI: 10.1038/s41598-018-25450-4.

Dobson SL., Bourtzis K., Braig HR., Jones BF., Zhou W., Rouset F., O’Neill SL. 1999. Wolbachia infections are distributed throughout insect somatic and germ line tissues. Insect biochemistry and molecular biology 29:153–60.

Dobson SL., Rattanadechakul W. 2001. A novel technique for removing Wolbachia infections from Aedes albopictus (Diptera: Culicidae). Journal of Medical Entomology 38:844–849.

Đorđević M., Stojković B., Savković U., Immonen E., Tucić N., Lazarević J., Arnvist G. 2017. Sex-specific mitonuclear epistasis and the evolution of mitochondrial bioenergetics, ageing, and life history in seed beetles. Evolution 71:274–288. DOI: 10.1111/evo.13109.

Dunning Hotopp JC., Clark ME., Oliveira DCSG., Foster JM., Fischer P., Torres MCM., Giebel JD., Kumar N., Ishmael N., Wang S., Ingram J., Nene R V., Shepard J., Tomkins J., Richards S., Spiro DJ., Ghedin E., Slatko BE., Tettelin H., Werren JH. 2007. Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. Science 317:1753–1756. DOI: 10.1126/science.1142490.

Duron O., Bouchon D., Boutin S., Bellamy L., Zhou L., Engelstader J., Hurst GD. 2008. The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. BMC Biology 6:27. DOI: 10.1186/1741-7007-6-27.

Ellegaard KM., Klasson L., Näslund K., Bourtzis K., Andersson SGE. 2013. Comparative Genomics of Wolbachia and the Bacterial Species Concept. PLoS Genetics 9:e1003381. DOI: 10.1371/journal.pgen.1003381.

Frost CL., Fernández-Marín H., Smith JE., Hughes WOH. 2010. Multiple gains and losses of Wolbachia symbionts across a tribe of fungus-growing ants. Molecular Ecology 19:4077–4085. DOI: 10.1111/j.1365-294X.2010.04764.x.

Gerth M., Gansauge M-T., Weigert A., Bleidorn C. 2014. Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic. Nature Communications 5. DOI:
Gerth M., Hurst GDD. 2017. Short reads from honey bee (Apis sp.) sequencing projects reflect microbial associate diversity. PeerJ 5:e3529. DOI: 10.7717/peerj.3529.

Ghedin E., Wang S., Foster JM., Slatko BE. 2004. First sequenced genome of a parasitic nematode. Trends in Parasitology 20:151–153. DOI: 10.1016/j.pt.2004.01.011.

Giordano R., Jackson JJ., Robertson HM. 1997. The role of Wolbachia bacteria in reproductive incompatibilities and hybrid zones of Diabrotica beetles and Gryllus crickets. Proceedings of the National Academy of Sciences of the United States of America 94:11439–44.

Glowska E., Dragun-Damian A., Dabert M., Gerth M. 2015. New Wolbachia supergroups detected in quill mites (Acari: Syringophilidae). Infection, Genetics and Evolution 30:140–146. DOI: 10.1016/J.MEEGID.2014.12.019.

Gomes FM., Hixson BL., Tyner MDW., Ramirez JL., Canepa GE., Alves e Silva TL., Molina-Cruz A., Keita M., Kane F., Traoré B., Sogoba N., Barillas-Mury C. 2017. Effect of naturally occurring Wolbachia in Anopheles gambiae s.l. mosquitoes from Mali on Plasmodium falciparum malaria transmission. Proceedings of the National Academy of Sciences of the United States of America 114:12566–12571. DOI: 10.1073/pnas.1716181114.

Griffin PC., Khadake J., LeMay KS., Lewis SE., Orchard S., Pask A., Pope B., Roessner U., Russell K., Seemann T., Treloar A., Tyagi S., Christiansen JH., Dayalan S., Gladman S., Hangartner SB., Hayden HL., Ho WWH., Keeble-Gagnère G., Korhonen PK., Neish P., Prestes PR., Richardson MF., Watson-Haigh NS., Wyres KL., Young ND., Schneider MV. 2017. Best practice data life cycle approaches for the life sciences. F1000Research 6:1618. DOI: 10.12688/f1000research.12344.1.

Gurevich A., Saveliev V., Vyahhi N., Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. DOI: 10.1093/bioinformatics/btt086.

Haegeman A., Vanholme B., Jacob J., Vandekerckhove TTM., Claeyts M., Borgenie G., Gheysen G. 2009. An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new Wolbachia supergroup. International Journal for Parasitology 39:1045–1054. DOI: 10.1016/J.IJPARA.2009.01.006.

Haine ER., Pickup NJ., Cook JM. 2005. Horizontal transmission of Wolbachia in a Drosophila community. Ecological Entomology 30:464–472. DOI: 10.1111/j.0307-6946.2005.00715.x.
Heath BD., Butcher RDJ., Whitfield WGF., Hubbard SF. 1999. Horizontal transfer of Wolbachia between phylogenetically distant insect species by a naturally occurring mechanism. *Current Biology* 9:313–316. DOI: 10.1016/S0960-9822(99)80139-0.

Hertig M. 1936. The Rickettsia, Wolbachia pipientis (gen. et sp.n.) and Associated Inclusions of the Mosquito, Culex pipiens. *Parasitology* 28:453. DOI: 10.1017/S0031182000022666.

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 Microbiology Letters* 281:215–220. DOI: 10.1111/j.1574-6968.2008.01110.x.

Hoffmann AA., Montgomery BL., Popovici J., Iturbe-Ormaetxe I., Johnson PH., Muzzi F., Greenfield M., Durkan M., Leong YS., Dong Y., Cook H., Axford J., Callahan AG., Kenny N., Omodei C., McGraw EA., Ryan PA., Ritchie SA., Turelli M., O’Neill SL. 2011. Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature* 476:454–457. DOI: 10.1038/nature10356.

Holden PR., Brookfield JFY., Jones P. 1993. Cloning and characterization of an ftsZ homologue from a bacterial symbiont of Drosophila melanogaster. *MGG Molecular & General Genetics* 240:213–220. DOI: 10.1007/BF00277059.

Holt RA. 2002. The Genome Sequence of the Malaria Mosquito Anopheles gambiae. *Science* 298:129–149. DOI: 10.1126/science.1076181.

Hughes GL., Koga R., Xue P., Fukatsu T., Rasgon JL. 2011. Wolbachia infections are virulent and inhibit the human malaria parasite Plasmodium falciparum in Anopheles gambiae. *PLoS Pathog* 7:e1002043. DOI: 10.1371/journal.ppat.1002043.

Hurst GDD., Jiggins FM., Hinrich Graf von der Schulenburg J., Bertrand D., West SA., Goriacheva II., Zakharov IA., Werren JH., Stouthamer R., Majerus MEN. 1999. Male-killing Wolbachia in two species of insect. *Proceedings of the Royal Society B: Biological Sciences* 266:735–740. DOI: 10.1098/rspb.1999.0698.

Husnik F., McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proceedings of the National Academy of Sciences* 113:E5416–E5424. DOI: 10.1073/pnas.1603910113.

Jamnongluk W., Kittayapong P., Baimai V., O’Neill SL. 2002. Wolbachia Infections of Tephritid Fruit Flies: Molecular Evidence for Five Distinct Strains in a Single Host Species. *Current Microbiology* 45:255–260. DOI: 10.1007/s00284-002-3746-1.
Jeffries CL., Lawrence GG., Golovko G., Kristan M., Orsborne J., Spence K., Hurn E., Bandibabone J., Tantely LM., Raharimalala FN., Keita K., Camara D., Barry Y., Watsenga F., Manzambi EZ., Afrane YA., Mohammed AR., Aboku TA., Hegde S., Khanipov K., Pimenova M., Fofanov Y., Boyer S., Irish SR., Hughes GL., Walker T. 2018. Novel Wolbachia strains in Anopheles malaria vectors from Sub-Saharan Africa. *bioRxiv*. DOI: 10.1101/338434.

Jeyaprakash A., Hoy MA. 2000. Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect molecular biology* 9:393–405.

Jiggins FM., Bentley JK., Majerus MEN., Hurst GDD. 2002. Recent changes in phenotype and patterns of host specialization in Wolbachia bacteria. *Molecular Ecology* 11:1275–1283. DOI: 10.1046/j.1365-294X.2002.01532.x.

Kageyama D., Ohno M., Sasaki T., Yoshido A., Konagaya T., Jouraku A., Kuwazaki S., Kanamori H., Katayose Y., Narita S., Miyata M., Riegler M., Sahara K. 2017. Feminizing Wolbachia endosymbiont disrupts maternal sex chromosome inheritance in a butterfly species. *Evolution Letters* 1:232–244. DOI: 10.1101/115386.

Katoh K., Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780. DOI: 10.1093/molbev/mst010.

Kautz S., Rubin BER., Moreau CS. 2013. Bacterial Infections across the Ants: Frequency and Prevalence of Wolbachia, Spiroplasma, and Asaia. *Psyche: A Journal of Entomology* 2013:1–11. DOI: 10.1155/2013/936341.

Kikuchi Y., Fukatsu T. 2003. Diversity of Wolbachia Endosymbionts in Heteropteran Bugs. *Applied and Environmental Microbiology* 69:6082–6090. DOI: 10.1128/AEM.69.10.6082-6090.2003.

Kittayapong P., Baisley KJ., Baimai V., O’Neill SL. 2000. Distribution and diversity of Wolbachia infections in Southeast Asian mosquitoes (Diptera: Culicidae). *Journal of medical entomology* 37:340–345. DOI: 10.1603/0022-2585(2000)037[0340:DADOWI]2.0.CO;2.

Kodandaramaiah U., Weingartner E., Janz N., Dalén L., Nylin S. 2011. Population structure in relation to host-plant ecology and Wolbachia infestation in the comma butterfly. *Journal of
Evolutionary Biology 24:2173–2185. DOI: 10.1111/j.1420-9101.2011.02352.x.

Kondo N., Ijichi N., Shimada M., Fukatsu T. 2002. Prevailing triple infection with Wolbachia in Callosobruchus chinensis (Coleoptera: Bruchidae). Molecular Ecology 11:167–180. DOI: 10.1046/j.0962-1083.2001.01432.x.

Kondo N., Shimada M., Fukatsu T. 1999. High Prevalence of Wolbachia in the Azuki Bean Beetle Callosobruchus chinensis (Coleoptera, Bruchidae). Zoological Science 16:955–962. DOI: 10.2108/zsj.16.955.

Koukou K., Pavlikaki H., Kilias G., Werren JH., Bourtzis K., Alahiotis SN. 2006. Influence of antibiotic treatment and Wolbachia curing on sexual isolation among Drosophila melanogaster cage populations. Evolution 60:87–96. DOI: 10.1111/j.0014-3820.2006.tb01084.x.

Kraaijeveld K., Franco P., de Knijff P., Stouthamer R., van Alphen JJ. 2011. Clonal genetic variation in a Wolbachia-infected asexual wasp: horizontal transmission or historical sex? Molecular Ecology 20:no-no. DOI: 10.1111/j.1365-294X.2011.05150.x.

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33:1870–1874. DOI: 10.1093/molbev/msw054.

Kunin V., Copeland A., Lapidus A., Mavromatis K., Hugenholtz P. 2008. A Bioinformatician’s Guide to Metagenomics. Microbiology and Molecular Biology Reviews 72:557–578. DOI: 10.1128/MMBR.00009-08.

Kyei-Poku GK., Giladi M., Coghlin P., Mokady O., Zchori-Fein E., Floate KD. 2006. Wolbachia in wasps parasitic on filth flies with emphasis on Spalangia cameroni. Entomologia Experimentalis et Applicata 121:123–135. DOI: 10.1111/j.1570-8703.2006.00469.x.

Leclercq S., Thézé J., Chebbi MA., Giraud I., Moumen B., Ernenwein L., Grève P., Gilbert C., Cordaux R. 2016. Birth of a W sex chromosome by horizontal transfer of Wolbachia bacterial symbiont genome. Proceedings of the National Academy of Sciences of the United States of America 113:15036–15041. DOI: 10.1073/pnas.1608979113.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Lindsey AR., Bordenstein SR., Newton IL., Rasgon JL. 2016a. Wolbachia pipientis should not be split into multiple species: A response to Ramírez-Puebla et al., “Species in Wolbachia? Proposal for the designation of ‘Candidatus Wolbachia bourtzisii’, ‘Candidatus Wolbachia
onchocercicola’, ‘Candidatus Wolbachia blax. *Systematic and Applied Microbiology*
39:220–222. DOI: 10.1016/j.syapm.2016.03.001.

Lindsey ARI., Werren JH., Richards S., Stouthamer R. 2016b. Comparative Genomics of a
Parthenogenesis-Inducing Wolbachia Symbiont. *G3: Genes-Genomes-Genetics* 6:2113–
2123. DOI: 10.1534/g3.116.028449.

Lo N., Casiraghi M., Salati E., Bazzocchi C., Bandi C. 2002. How Many Wolbachia Supergroups
Exist? *Molecular Biology and Evolution* 19:341–346. DOI:
10.1093/oxfordjournals.molbev.a004087.

Lo N., Paraskevopoulos C., Bourtzis K., O’Neill SL., Werren JH., Bordenstein SR., Bandi C.
2007. Taxonomic status of the intracellular bacterium Wolbachia pipientis. *International
Journal of Systematic and Evolutionary Microbiology* 57:654–657. DOI:
10.1099/IJS.0.64515-0.

Maier CT. 1981. Parasitoids Emerging from Puparia of Rhagoletis Pomonella (Diptera:
Tephritidae) Infesting Hawthorn and Apple in Connecticut. *The Canadian Entomologist*
113:867–870. DOI: 10.4039/Ent113867-9.

Masui S., Sasaki T., Ishikawa H. 1997. groE-Homologous Operon of Wolbachia, an Intracellular
Symbiont of Arthropods: A New Approach for Their Phylogeny. *Zoological Science*
14:701–706. DOI: 10.2108/zsj.14.701.

Mateos M., Castrezana SJ., Nankivell BJ., Estes AM., Markow TA., Moran NA. 2006. Heritable
endosymbionts of Drosophila. *Genetics* 174:363–376. DOI: 10.1534/genetics.106.058818.

Mock A., Tajovsky K., Žurovcová M., Jarošová A., Kocourek P., Gruber J., Angyal D., Spelda J.
2016. Hungarosoma bokori Verhoeff, 1928 (Diplopoda: Chordeumatida): new insights into
its taxonomy, systematics, molecular genetics, biogeography and ecology. *Zootaxa*
4178:234. DOI: 10.11646/zootaxa.4178.2.4.

Narita S., Nomura M., Kageyama D. 2007. Naturally occurring single and double infection with
Wolbachia strains in the butterfly *Eurema hecabe*: transmission efficiencies and population
density dynamics of each Wolbachia strain. *FEMS Microbiology Ecology* 61:235–245.
DOI: 10.1111/j.1574-6941.2007.00333.x.

O’Neill SL., Giordano R., Colbert AM., Karr TL., Robertson HM. 1992. 16S rRNA phylogenetic
analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in
insects. *Proceedings of the National Academy of Sciences of the United States of America*
Ødegaard F. 2000. How many species of arthropods? Erwin’s estimate revised. *Biological Journal of the Linnean Society* 71:583–597. DOI: 10.1006/BIJL.2000.0468.

Osborne SE., Leong YS., O’Neill SL., Johnson KN. 2009. Variation in Antiviral Protection Mediated by Different Wolbachia Strains in Drosophila simulans. *PLoS Pathogens* 5:e1000656. DOI: 10.1371/journal.ppat.1000656.

Paraskevopoulos C., Bordenstein SR., Wernegreen JJ., Werren JH., Bourtzis K. 2006. Toward a Wolbachia multilocus sequence typing system: Discrimination of Wolbachia strains present in Drosophila species. *Current Microbiology* 53:388–395. DOI: 10.1007/s00284-006-0054-1.

Perrot-Minnot MJ., Guo LR., Werren JH. 1996. Single and double infections with Wolbachia in the parasitic wasp Nasonia vitripennis: effects on compatibility. *Genetics* 143:961–72.

Pintureau B., Chaudier S., Lassablière F., Charles H., Grenier S. 2000. Addition of wsp sequences to the Wolbachia phylogenetic tree and stability of the classification. *Journal of Molecular Evolution* 51:374–377. DOI: 10.1007/s002390010099.

Plantard O., Rasplus JY., Mondor G., Le Clainche I., Solignac M. 1999. Distribution and phylogeny of Wolbachia inducing thelytoky in Rhoditini and “Aylacini” (Hymenoptera: Cynipidae). *Insect Molecular Biology* 8:185–91.

Quinlan AR., Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.

Ramirez-Puebla ST., Ormeno-Orrillo E., Vera-Ponce de Leon A., Lozano L., Sanchez A., Rosenblueth M., Martinez-Romero E. 2016. Genomes of Candidatus Wolbachia Bourtzisii wDacA and Candidatus Wolbachia Pipientis wDacB from the Cochineal Insect Dactylopius coccus (Hemiptera: Dactylopiidae). *G3: Genes-Genomes-Genetics* 6:3343–3349. DOI: 10.1534/g3.116.031237.

Rasgon JL., Scott TW. 2004. An Initial Survey for Wolbachia (Rickettsiales: Rickettsiaceae) Infections in Selected California Mosquitoes (Diptera: Culicidae) : Table 1. *Journal of Medical Entomology* 41:255–257. DOI: 10.1603/0022-2585-41.2.255.

Raychoudhury R., Baldo L., Oliveira DCSG., Werren JH. 2009. Modes of acquisition of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the Nasonia species complex. *Evolution* 63:165–183. DOI: 10.1111/j.1558-5646.2008.00533.x.
Ricci I., Cancrini G., Gabrielli S., D’amelio S., Favia G. 2002. Searching for Wolbachia (Rickettsiales: Rickettsiaceae) in Mosquitoes (Diptera: Culicidae): Large Polymerase Chain Reaction Survey and New Identifications. *Journal of Medical Entomology* 39:562–567. DOI: 10.1603/0022-2585-39.4.562.

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 Genetics* 8. DOI: 10.1371/journal.pgen.1003129.

Riegler M., Charlat S., Stauffer C., Merçot H. 2004. Wolbachia Transfer from Rhagoletis cerasi to Drosophila simulans: Investigating the Outcomes of Host-Symbiont Coevolution. *Applied and Environmental Microbiology* 70:273–279. DOI: 10.1128/AEM.70.1.273–279.2004.

Rogers EE., Backus EA. 2014. Anterior foregut microbiota of the glassy-winged sharpshooter explored using deep 16S rRNA gene sequencing from individual insects. *PLoS ONE* 9. DOI: 10.1371/journal.pone.0106215.

Rokas A., Atkinson RJ., Brown GS., West SA., Stone GN. 2001. Understanding patterns of genetic diversity in the oak gallwasp Biorhiza pallida: demographic history or a Wolbachia selective sweep? *Heredity* 87:294–304. DOI: 10.1046/j.1365-2540.2001.00872.x.

Rokas A., Atkinson RJ., Nieves-Aldrey JL., West SA., Stone GN. 2002. The incidence and diversity of Wolbachia in gallwasps (Hymenoptera; Cynipidae) on oak. *Molecular Ecology* 11:1815–1829. DOI: 10.1046/j.1365-294X.2002.01556.x.

Ros VID., Fleming VM., Feil EJ., Breeuwer JAJ. 2009. How diverse is the genus Wolbachia? Multiple-gene sequencing reveals a putatively new Wolbachia supergroup recovered from spider mites (Acari: Tetranychidae). *Applied and Environmental Microbiology* 75:1036–43. DOI: 10.1128/AEM.01109-08.

Ros VID., Fleming VM., Feil EJ., Breeuwer JAJ. 2012. Diversity and recombination in Wolbachia and Cardinium from Bryobia spider mites. *BMC Microbiology* 12. DOI: 10.1186/1471-2180-12-S1-S13.

Rousset F., Braig HR., O’Neill SL. 1999. A stable triple Wolbachia infection in Drosophila with nearly additive incompatibility effects. *Heredity* 82 (Pt. 6):620–7.

Rowley SM., Raven RJ., McGraw EA. 2004. Wolbachia pipiensis in Australian Spiders. *Current Microbiology* 49. DOI: 10.1007/s00284-004-4346-z.
Russell JA., Funaro CF., Giraldo YM., Goldman-Huertas B., Suh D., Kronauer DJC., Moreau CS., Pierce NE. 2012. A Veritable Menagerie of Heritable Bacteria from Ants, Butterflies, and Beyond: Broad Molecular Surveys and a Systematic Review. *PLoS ONE* 7:e51027. DOI: 10.1371/journal.pone.0051027.

Saha S., Hunter WB., Reese J., Morgan JK., Marutani-Hert M., Huang H., Lindeberg M. 2012. Survey of Endosymbionts in the *Diaphorina citri* Metagenome and Assembly of a *Wolbachia* wDi Draft Genome. *PLoS ONE* 7:e50067.

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 Biology* 6. DOI: 10.1186/gb-2005-6-3-r23.

Saridaki A., Bourtzis K. 2010. Wolbachia: more than just a bug in insects genitals. *Current Opinion in Microbiology* 13:67–72. DOI: 10.1016/j.mib.2009.11.005.

Schilthuizen M., Honda J., Stouthamer R. 1998. Parthenogenesis-inducing Wolbachia in Trichogramma kaykai (hymenoptera: Trichogrammatidae) originates from a single infection. *Annals of the Entomological Society of America* 91:410–414. DOI: 10.1093/aesa/91.4.410.

Schilthuizen M., Stouthamer R. 1998. Distribution of Wolbachia among the guild associated with the parthenogenetic gall wasp Diplolepis rosae. *Heredity* 81:270–274. DOI: 10.1038/sj.hdy.6883850.

Schmieder R., Edwards R. 2011. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS ONE* 6:e17288. DOI: 10.1371/journal.pone.0017288.

Schuler H., Arthofer W., Riegler M., Bertheau C., Krumböck S., Köppler K., Vogt H., Teixeira LAF., Stauffer C. 2011. Multiple Wolbachia infections in Rhagoletis pomonella. *Entomologia Experimentalis et Applicata* 139:138–144. DOI: 10.1111/j.1570-7458.2011.01115.x.

Simão FA., Waterhouse RM., Ioannidis P., Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. DOI: 10.1093/bioinformatics/btv351.

Simões PM., Mialdea G., Reiss D., Sagot M-F., Charlat S. 2011. Wolbachia detection: an assessment of standard PCR Protocols. *Molecular Ecology Resources* 11:567–572. DOI:
Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.

Stelinski LL., Pelz KS., Liburd OE. 2004. Field Observations Quantifying Attraction of the Parasitic Wasp, Diachasma Alloeum (Hymenoptera: Braconidae) To Blueberry Fruit Infested By the Blueberry Maggot Fly, Rhagoletis Mendax (Diptera: Tephritidae). *Florida Entomologist* 87:124–129. DOI: 10.1653/0015-4040(2004)087[0124:FOQAOT]2.0.CO;2.

Stouthamer R., Luck RF., Hamilton WD. 1990. Antibiotics cause parthenogenetic Trichogramma (Hymenoptera/Trichogrammatidae) to revert to sex. *Proceedings of the National Academy of Sciences of the United States of America* 87:2424–7. DOI: 10.1073/pnas.87.7.2424.

Subandiyah S., Nikoh N., Tsuyumu S., Somowiyarjo S., Fukatsu T. 2000. Complex Endosymbiotic Microbiota of the Citrus Psyllid Diaphorina citri (Homoptera: Psylloidea). *Zoological Science* 17:983–989. DOI: 10.2108/zsj.17.983.

Teixeira L., Ferreira Á., Ashburner M. 2008. The Bacterial Symbiont Wolbachia Induces Resistance to RNA Viral Infections in Drosophila melanogaster. *PLoS Biology* 6:e1000002. DOI: 10.1371/journal.pbio.1000002.

Tsai K-H., Lien J-C., Huang C-G., Wu W-J., Chen W-J. 2004. Molecular (sub) grouping of endosymbiont Wolbachia infection among mosquitoes of Taiwan. *Journal of medical entomology* 41:677–83. DOI: Doi 10.1603/0022-2585-41.4.677.

Vavre F., Fleury F., Lepetit D., Fouillet P., Boulétreau M. 1999. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. *Molecular Biology and Evolution* 16:1711–23.

Verne S., Johnson M., Bouchon D., Grandjean F. 2007. Evidence for recombination between feminizing Wolbachia in the isopod genus Armadillidium. *Gene* 397:58–66. DOI: 10.1016/j.gene.2007.04.006.

Walker T., Moreira LA. 2011. Can Wolbachia be used to control malaria? *Memórias do Instituto Oswaldo Cruz* 106:212–217. DOI: 10.1590/S0074-02762011000900026.

Wang Y., Chandler C. 2016. Candidate pathogenicity islands in the genome of “Candidatus
Rickettsiella isopodorum”, an intracellular bacterium infecting terrestrial isopod crustaceans. PeerJ 2016. DOI: 10.7717/peerj.2806.

Weeks AR., Turelli M., Harcombe WR., Reynolds KT., Hoffmann AA. 2007. From Parasite to Mutualist: Rapid Evolution of Wolbachia in Natural Populations of Drosophila. PLoS Biology 5:e114. DOI: 10.1371/journal.pbio.0050114.

Werren JH. 1997a. Wolbachia run amok. Proceedings of the National Academy of Sciences of the United States of America 94:11154–11155.

Werren JH. 1997b. Biology of Wolbachia. Annual Review of Entomology 42:587–609. DOI: 10.1146/annurev.ento.42.1.587.

Werren JH., Bartos JD. 2001. Recombination in Wolbachia. Current Biology 11:431–435. DOI: 10.1016/S0960-9822(01)00101-4.

Werren JH., Windsor DM. 2000. Wolbachia infection frequencies in insects: evidence of a global equilibrium? Proceedings of the Royal Society B: Biological Sciences 267:1277–85. DOI: 10.1098/rspb.2000.1139.

Werren JH., Windsor D., Guo L. 1995. Distribution of Wolbachia among Neotropical Arthropods. Proceedings of the Royal Society B: Biological Sciences 262:197–204. DOI: 10.1098/rspb.1995.0196.

Werren JH., Zhang W., Guo LR. 1995. Evolution and Phylogeny of Wolbachia: Reproductive Parasites of Arthropods. Proceedings of the Royal Society B: Biological Sciences 261:55–63. DOI: 10.1098/rspb.1995.0117.

Westermann AJ., Gorski SA., Vogel J. 2012. Dual RNA-seq of pathogen and host. Nature Reviews Microbiology 10:618–630. DOI: 10.1038/nrmicro2852.

Wilkinson MD., Dumontier M., and others. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3. DOI: doi:10.1038/sdata.2016.18.

Wiwatanaratanabutr I., Kittayapong P., Caubet Y., Bouchon D. 2009. Molecular Phylogeny of Wolbachia Strains in Arthropod Hosts Based on groE-Homologous Gene Sequences. Zoological Science 26:171–177. DOI: 10.2108/zsj.26.171.

Zerbino DR., Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18:821–829. DOI: 10.1101/gr.074492.107.

Zha X., Zhang W., Zhou C., Zhang L., Xiang Z., Xia Q. 2014. Detection and characterization of Wolbachia infection in silkworm. Genetics and Molecular Biology 37:573–580. DOI:
Zhou W., Rousset F., O’Neil S. 1998. Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. *Proceedings of the Royal Society B: Biological Sciences* 265:509–15. DOI: 10.1098/rspb.1998.0324.
Concatenated phylogeny

Molecular phylogenetic analysis by maximum likelihood based on the concatenated dataset containing *ftsZ* and the *groE* operon (total of 1381 nucleotide positions). Bold letters next to host species names indicate supergroup relationships of *Wolbachia* isolates identified in previous studies for each host. Asterisks indicate reference sequences (see also Table 2).
Figure 2 (on next page)

Whole-genome phylogeny

Maximum likelihood phylogeny based on whole-genome sequence data of *Wolbachia* isolates assembled here and previously sequenced reference *Wolbachia* genomes (indicated by samples with associated accession numbers), with a total of 133,744 nucleotide positions. Numbers by nodes indicate bootstrap support based on 200 replicates. Bold letters next to host species names indicate supergroup relationships of *Wolbachia* isolates identified in previous studies for each host. Isolates with accession numbers listed represent reference genome sequences from other studies.
Table 1 (on next page)

Reference Wolbachia genes

Gene sequences from *Wolbachia*-infected hosts used to create the reference database for Magic-BLAST searches of SRA accessions to diagnose novel *Wolbachia* infections.
Table 1. Gene sequences from *Wolbachia*-infected hosts used to create the reference database for Magic-BLAST searches of SRA accessions to diagnose novel *Wolbachia* infections.

| Host Classification | Phylum      | Class    | Order   | Species        | Supergroup | Strain     | Accession Number | Gene | Citation                          |
|---------------------|-------------|----------|---------|----------------|------------|------------|------------------|------|-----------------------------------|
|                     | Arachnida   | Trombidiformes |        | Bryobia praetiosa | B         | —         | JN572870.1       | wsp  | Ros et al. 2012                   |
|                     |             |          |         | *Diaea circumlita* | A         | wDiacri3  | AY486091.1       | wsp  | Rowley et al. 2004                |
|                     | Coleoptera  |          |         | *Tribolium confusum* | —         | NFRI114    | AB469356.1       | wsp  | Kageyama et al. 2008              |
|                     |             |          |         | *Tribolium confusum* | —         | —         | DQ842337.1       | ftsZ | Baldo et al. 2006                 |
|                     |             |          |         | *Dicladispa armigera* | A         | wDic      | DQ243935.1       | groE | Wiwatanaratnamburt et al. 2009    |
|                     |             |          |         | *Culex pipiens* | B         | —         | DQ900650.1       | wsp  | Djadid and Zakeri 2006 (unpublished) |
|                     |             |          |         | *Drosophila melanogaster* | A         | wMel      | FJ403332.1       | wsp  | Wang and Zheng 2008 (unpublished)  |
|                     |             |          |         | *Drosophila simulans* | A         | wMel      | DQ412101.1       | wsp  | Mateos et al. 2006                |
|                     |             |          |         | *Protosiphon stali* | B         | wProtPA   | AF448376.1       | wsp  | Werren and Bartos 2001            |
|                     |             |          |         | *Culex quinquefasciatus* | B         | 22        | GU901159.1       | ftsZ | Rondun-Duenas et al. 2010 (unpublished) |
|                     |             |          |         | *Drosophila melanogaster* | A         | Canton-S  | X71906.1         | ftsZ | Holdren et al. 1993               |
|                     |             |          |         | *Drosophila recens* | A         | —         | U28174.1         | ftsZ | Werren et al. 1995                |
|                     |             |          |         | *Drosophila simulans* | A         | wHa       | AY508996.1       | ftsZ | Ballard 2003 (unpublished)         |
|                     |             |          |         | *Drosophila simulans* | A         | wMa(Ma)   | AY508999.1       | ftsZ | Ballard 2003 (unpublished)         |
|                     |             |          |         | *Aedes albopictus* | A         | wAlbA     | DQ243927.1       | groE | Wiwatanaratnamburt et al. 2009    |
|                     |             |          |         | *Culex fasciiculatus* | B         | wFusc     | AJ511284.1       | groE | Wiwatanaratnamburt and Kittayapong 2002 (unpublished) |
|                     |             |          |         | *Drosophila simulans* | A         | —         | AB002287.1       | groE | Masui et al. 1997                |
|                     |             |          |         | *Drosophila simulans* | A         | —         | AB002288.1       | groE | Masui et al. 1997                |
|                     |             |          |         | *Drosophila trietus* | A         | —         | AY563553.1       | groE | Haine et al. 2005                |
|                     |             |          |         | *Encarsia formosa* | B         | —         | KC161951.1       | wsp  | Lu and Jiang 2012 (unpublished)    |
|                     |             |          |         | *Muscidifuraxuriaraptor* | A         | —         | DQ380857.1       | wsp  | Kyei-Poku et al. 2006             |
|                     |             |          |         | *Nasonia girardi*   | —         | wNGirVA   | AF448381.1       | wsp  | Werren and Bartos 2001            |
|                     | Hymenoptera |          |         |                |            |           |                  |      |                                   |
| Family               | Species                  | Genus       | Species (nr. deion) | Genus       | Species (nr. deion) | Genus       | Species (nr. deion) | Reference                  |
|----------------------|--------------------------|-------------|---------------------|-------------|---------------------|-------------|---------------------|----------------------------|
| Lepidoptera          | Nasonia vitripennis      | —           | wNvi-2              | KC161919.1  | wsp                 | Lu and Jiang 2012  |
|                      | Trichogramma cordobensis | B           | Sib                 | AF245164.1  | wsp                 | Pintureau et al. 2000 |
|                      | Diplosoma rosae          | B           | Type I              | U83887.1    | ftsZ                | Schilthuizen and Stouthamer 1998 |
|                      | Habrocytus bedegerarii   | B           | Type I              | U83886.1    | ftsZ                | Schilthuizen and Stouthamer 1998 |
|                      | Trichogramma n. spec. (nr. deion) | B           | —                   | U59696.1    | ftsZ                | Schilthuizen and Stouthamer 1998 |
|                      | Asobara tabida           | —           | —                   | AJ634749.1  | groE                | Haine et al. 2005 |
|                      | Bombyx mandarina         | —           | —                   | KJ65909.1   | wsp                 | Zha et al. 2014 |
| Orthoptera           | Acraea encedon           | B           | —                   | AJ138892.1  | ftsZ                | Hurst et al. 1999 |
|                      | Ephesia kuehniella       | A           | Type II             | U62126.1    | ftsZ                | Schilthuizen et al. 1998 |
| Malacostraca         | Acrasia pharalis         | B           | —                   | AJ318481.1  | groE                | Jiggins et al. 2002 |
|                      | Ephesia cautella         | A           | —                   | AB002289.1  | groE                | Musui et al. 1997 |
|                      | Ephesia cautella         | B           | —                   | AB002290.1  | groE                | Musui et al. 1997 |
| Nematoda             | Gyrillos pennsylvanicus  | B           | —                   | U28195.1    | ftsZ                | Werren et al. 1995 |
| Isopoda              | Chaetophiloscia elongata | —           | —                   | AM887239.1  | groE                | Wivataranaranabat et al. 2009 |
| Secernentea          | Dirofilaria repens       | C           | —                   | AJ252176.1  | wsp                 | Bazzocchi et al. 2000 |
| Spiruidea            | Litomosoides sigmodontis | D           | —                   | AJ252177.1  | wsp                 | Bazzocchi et al. 2000 |
|                      | Onchocerca gibsoni       | C           | —                   | AJ252178.1  | wsp                 | Bazzocchi et al. 2000 |
|                      | Brugia malayi            | D           | —                   | AJ252061.1  | wsp                 | Bazzocchi et al. 2000 |
|                      | Brugia pahangi           | D           | —                   | AJ252175.1  | wsp                 | Bazzocchi et al. 2000 |
|                      | Dirofilaria immitis       | C           | —                   | AJ252062.1  | wsp                 | Bazzocchi et al. 2000 |
|                      | Wuchereria bancrofti     | D           | —                   | AF285273.1  | groE                | Salahuddeen and Nutman 2000  |
Table 2 (on next page)

Wolbachia sequences of known origin for phylogenetic analysis

Wolbachia genes used as controls and the species name from which they were isolated. The supergroup of the Wolbachia strain is listed and these genes served as a control during the creation of the phylogeny.
Table 2. Wolbachia genes used as controls and the species name from which they were isolated. The supergroup of the *Wolbachia* strain is listed and these genes served as a control during the creation of the phylogeny.

| Host Classification | Gene & Accession Number | Supergroup | Citation              |
|---------------------|--------------------------|------------|-----------------------|
| Phylum              |                          |            |                       |
| Arthropoda          |                          |            |                       |
| Isopoda             | Armidillidium vulgare    | DQ778101   | DQ778104 | B | Verne et al., 2007 |
| Hemiptera           | *Bemisia afer*           | KF452573 | KF452533 | B | Bing et al., 2014 |
| Hemiptera           | *Bemisia tabaci*         | KF452577 | KF452536 | B | Bing et al., 2014 |
| Diptera             | *Drosophila ambigua*     | AY563550  | AY563552 | A | Haine et al., 2005 |
| Diptera             | *Drosophila melanogaster*| DQ235339 | DQ235379 | A | Paraskevopoulos et al., 2006 |
| Diptera             | *Drosophila tristis*     | AY563551  | AY563553 | A | Haine et al., 2005 |
Table 3 (on next page)

Species showing evidence of Wolbachia infection

List of unique species (class; order) that tested positive for the presence of *Wolbachia* genes. Species indicated with a ‘*’ are species that have not previously been identified, according to our literature search, to harbor *Wolbachia* strains. The supergroup classification of the *Wolbachia* strain according to this study and previously studies is listed if known.
Table 3. List of unique species (class; order) that tested positive for the presence of Wolbachia genes. Species indicated with a ‘**’ are species that have not previously been identified, according to our literature search, to harbor Wolbachia strains. The supergroup classification of the Wolbachia strain according to this study and previously studies is listed if known.

| Phylum          | Class   | Order      | Species                  | Supergroup (this study) | Supergroup (previous study) | Citation                                                                 | GenBank Accession Numbers                          |
|-----------------|---------|------------|--------------------------|-------------------------|-----------------------------|--------------------------------------------------------------------------|---------------------------------------------------|
| Arthropoda      | Insecta | Diptera    | * Bembidion lapponicum   | B                       | —                           | —                                                                        | —                                                 |
|                 |         |            | Callosobruchus chinensis | A (but possible double infection) | A/B                         | Kondo et al., 1999; Kondo et al., 2002b; Kageyama et al., (unpublished) | AB025965; AB080664, AB080665, and AB081842 ; AB469358 |
|                 | Coleoptera |            | Diabrotica virgifera virgifera | A                       | A                           | Giordano et al., 1997; Giordano et al., (unpublished data)               | U83098, AF011270 - AF011271; DQ091306 - DQ091309 |
|                 |         |            | * Diploeciton nevermanni  | B                       | —                           | —                                                                        | —                                                 |
|                 |         |            | * Ectophya simulans       | B                       | —                           | —                                                                        | —                                                 |
|                 |         |            | Anopheles gambiae         | B                       | A                           | Baldini et al., 2014                                                    | KJ728739 - KJ728755                               |
|                 |         |            | Drosophila melanogaster   | A                       | A                           | Bourtzis et al., 1994; Woolfit et al., (unpublished data)               | Z28981 - Z28983; KJ440871 - KJ440895              |
|                 |         |            | Drosophila simulans       | A/B                     | A/B                         | Riegler et al., 2004; Ellegaard et al., 2013                            | AY227739, AY227742; CP003883, CP003884            |
|                 |         |            | Drosophila triauraria     | A                       | A                           | Cordaux et al., 2008                                                   | EU714523                                          |
|                 |         |            | Drosophila yakuba         | A                       | A                           | Charlat et al., 2004; Zabalou et al., 2004; Ioannidis et al., 2007; Cordaux et al., 2008 | AY291346, AY291348; AJ620679; DQ498875; EU714519 |
|                 |         |            | Rhagoletis pomonella      | A                       | A                           | Schuler et al., 2011                                                   | HQ333145 - HQ333159                               |
|                 |         |            | Rhagoletis zephyria       | A                       | A                           | Schuler et al., 2011                                                   | —                                                 |
|                 |         |            | Dactylopius coccus        | B                       | A/B                         | Ramírez-Puebla et al., 2016                                           | LSYX000000000, LSYY000000000                       |
|                 |         |            | Diaphorina citri          | B                       | B                           | Subandiyah et al., 2000; Lindsey et al., 2016b                         | AB038366-AB038370                                 |
|                 | Hymenoptera |            | * Gerris bueni             | B                       | —                           | —                                                                        | —                                                 |
|                 |         |            | Homalodisca vitripennis   | B                       | B                           | Rogers & Backus, 2014                                                  | KF636751                                          |
|                 |         |            | Maconellicoccus hirsutus   | B                       | B                           | Husnik & McCutcheon, 2016                                              | PRJEB12066 (European Nucleotide Archive)          |
|                      | Scientific Name               | Publication | Accession Numbers                                                                 |
|----------------------|------------------------------|-------------|-----------------------------------------------------------------------------------|
| Hymenoptera          | Megacopta cribraria          | A           | Kikuchi & Fukatsu, 2003; Jenkins et al., (unpublished data); AB109601, AB109602, JQ266093 |
|                      | Mycopsylla fici              | —           | Fromont et al., (unpublished data); KT273254, KT273255, KT273261, KT273277        |
|                      | Mycopsylla proxima           | —           | Fromont et al., (unpublished data); KT273257, KT273259, KT273260, KT273278        |
|                      | Acromyrmex echinatior        | A           | Frost et al., 2010; HM211007 - HM211071                                           |
|                      | Biorhiza pallida             | A           | Rokas et al., 2001; AF339629                                                     |
|                      | * Ceratina calcarata         | —           | —                                                                                  |
|                      | Cynipini sp.                 | A           | Abe & Miura, 2002; AB052667                                                     |
|                      | * Diacrasma allosum          | A           | —                                                                                  |
|                      | Diplolepis spinosa           | B           | Plantard et al., 1999; AF034987                                                  |
|                      | * Isocolus centaureae        | B           | —                                                                                  |
|                      | Pediaspis aceris             | A           | Rokas et al., 2002; —                                                            |
|                      | Pseudomyrmex sp. PSW-54      | A           | Kautz et al., 2013; KF015789                                                     |
|                      | Trichogramma pretiosum       | B           | Lindsey et al., 2016; LKEQ00000000                                               |
| Lepidoptera          | * Delias oria                | B           | —                                                                                  |
|                      | Operophtera brumata          | B           | Derks et al., 2015; JYPC00000000                                                 |
|                      | Pararge aegeria              | B           | Russell et al., 2012; KC137224                                                   |
|                      | Polygonia c-album            | B           | Kodandaramaiah et al., 2011; JN093149 - JN093153                                |
Table 4 (on next page)

Wolbachia genome assemblies

Information on Wolbachia draft genome assemblies. Expected host coverage is calculated as (total sequence data / host genome size). "Evidence of multiple infections" indicates whether or not the assembly contains signs pointing to multiple, distinct Wolbachia strains within the same biological host sample used for generating the sequence data (though some of these consisted of pooled individuals). BUSCO comp., BUSCO dup., BUSCO frag., and BUSCO missing refer to the number of BUSCO orthologs that were found to be complete and single copy, duplicated, fragmented, and missing from the Wolbachia assembly, out of 148 BUSCOs present in the Bacteria odb9 reference gene set. Grey rows at the bottom of the table were omitted from the whole-genome phylogenetic analysis because the assemblies appeared less complete (as indicated by missing BUSCO genes) or showed evidence of being chimeric or a mixture of two independent strains.
Table 4. Information on *Wolbachia* draft genome assemblies. Expected host coverage is calculated as (total sequence data / host genome size). "Evidence of multiple infections" indicates whether or not the assembly contains signs pointing to multiple, distinct *Wolbachia* strains within the same biological host sample used for generating the sequence data (though some of these consisted of pooled individuals). BUSCO comp., BUSCO dup., BUSCO frag., and BUSCO missing refer to the number of BUSCO orthologs that were found to be complete and single copy, duplicated, fragmented, and missing from the *Wolbachia* assembly, out of 148 BUSCOs present in the Bacteria *odb9* reference gene set. Grey rows at the bottom of the table were omitted from the whole-genome phylogenetic analysis because the assemblies appeared less complete (as indicated by missing BUSCO genes) or showed evidence of being chimeric or a mixture of two independent strains.

| Host species/ID | Description/ common name | BioSample accession number | SRA accession numbers | Total seq. data (Gb) | Host genome size (ref.) | Expected host coverage (x) | *Wolbachia* median cov. (x) | *Wolbachia* assembly size (Mbp) | *Wolbachia* assembly N50 (kb) | BUSCO comp. | BUSCO dup. | BUSCO frag. | BUSCO missing | Evidence of multiple infections? | Sample notes |
|----------------|--------------------------|----------------------------|-----------------------|----------------------|--------------------------|---------------------------|-----------------------------|-----------------------------|------------------------|----------------|----------------|----------------|----------------|--------------------------------|--------------|
| Anopheles gambiae | Mosquito | SAMEA3911293 | ERR1554906 ERR1554870 ERR1554834 | 9.1 | 280 Mb (Holt 2002) | 32 | 9.0 | 1.212 | 23.12 | 125 | 0 | 2 | 21 | No |
| Biorhiza pallida 1 | Wasp | SAMEA2053316 | ERR233308 | 8.7 | 16 | 1.249 | 9.5 | 125 | 0 | 5 | 18 | No |
| Biorhiza pallida 2 | Wasp | SAMEA2053315 | ERR233309 | 8.3 | 16 | 1.246 | 10.29 | 128 | 0 | 3 | 17 | No |
| Delias orasius | Butterfly | SAMN05712507 | SRR4341246 | 13.2 | 26 | 1.207 | 13.35 | 124 | 0 | 4 | 20 | No |
| Diabrotica virgifera virgifera 1 | Western corn rootworm | SAMN02373824 | SRR1106898 SRR1106897 SRR1106544 | 98.9 | 2.4 Gb (GCA_0003013835.1) | 41 | 900 | 1.505 | 31 | 128 | 0 | 3 | 17 | No |
| Diabrotica virgifera virgifera 2 | Western corn rootworm | SAMN02373827 | SRR1106912 SRR1106546 | 95.1 | 2.4 Gb (GCA_0003013835.1) | 40 | 750 | 1.487 | 35.01 | 128 | 0 | 3 | 17 | No |
| Diabrotica virgifera virgifera 3 | Western corn rootworm | SAMN02373842 | SRR1107707 SRR1107708 SRR1107710 SRR1107712 | 88.9 | 2.4 Gb (GCA_0003013835.1) | 37 | 660 | 1.376 | 32.16 | 128 | 0 | 3 | 17 | No |
| Diachasma alloeum | Wasp | SAMN03701895 | SRR2042503 SRR2046752 | 56.8 | 390 Mb (GCA_00141251.1) | 150 | 830 | 1.377 | 21.39 | 127 | 1 | 2 | 18 | No |
| Diaphorina citri 1 | Asian citrus psyllid | SAMN00100712 | SRR189238 SRR183690 | 25 | 490 Mb (GCA_000475195.1) | 51 | 180 | 1.379 | 25.7 | 128 | 0 | 2 | 18 | No |
| Diaphorina citri 2 | Asian citrus psyllid | SAMN01886038 | SRR649417 SRR649429 SRR649431 SRR649432 SRR649434 | 27.1 | 490 Mb (GCA_000475195.1) | 55 | 250 | 1.425 | 25.7 | 128 | 0 | 2 | 18 | No |
| Diploceiton nevermanni | Beetle | SAMN05860871 | SRR4342174 | 24.8 | 63 | 1.698 | 10.59 | 124 | 0 | 5 | 19 | No |
| Species                        | Type              | GenBank ID   | ERR ID(s)                  | Assembly Size | Nucleosome Size | Estimated ploidy | Population | Sex            | Comments                                |
|-------------------------------|-------------------|--------------|----------------------------|---------------|----------------|------------------|------------|----------------|-----------------------------------------|
| Diplolepis spinosa 1          | Gall wasp         | SAMEA3930570 | ERR1359308                 | 6.8           | 57             | 1.398            | 12.93      | 127            | 0 3 18 No                               |
| Diplolepis spinosa 2          | Gall wasp         | SAMEA3930574 | ERR1359312                 | 7.1           | 53             | 1.382            | 12.13      | 127            | 0 3 24 No                               |
| Drosophila melanogaster 1     | Fruit fly         | SAMEA3634594 | ERR1092813 ERR1092814 ERR1092815 ERR1092816 ERR1092817 ERR1092818 | 18.1 175 Mb 100 1600 1208 | 19.19 | 127            | 0 3 18 No                               |
| Drosophila melanogaster 2     | Fruit fly         | SAMN04017483 | SRR2347338                 | 3.5 175 Mb 20 21 1.198 | 13.83 | 125            | 0 4 19 No                               |
| Drosophila simulans 1         | Fruit fly         | SAMEA4393562 | ERR1597896                 | 25.3 150 Mb 170 390 1.265 | 14.67 | 130            | 0 2 16 No                               |
| Drosophila simulans 2         | Fruit fly         | SAMEA4394322 | ERR1597899                 | 23.4 150 Mb 160 1100 1.294 | 15.57 | 130            | 0 2 16 No                               |
| Drosophila simulans 3         | Fruit fly         | SAMEA4394323 | ERR1597900                 | 37 150 Mb 250 1600 1.313 | 15.72 | 130            | 0 2 16 No                               |
| Drosophila triaurearia 1      | Fruit fly         | SAMD00051863 | DRR061000                  | 22.6 170 Mb 130 150 1.306 | 22.94 | 129            | 0 3 16 No                               |
| Drosophila yakuba 1           | Fruit fly         | SAMN04044077 | SRR2318687                 | 8 170 Mb 47 42 1.254 | 14.75 | 128            | 0 2 18 No                               |
| Drosophila yakuba 2           | Fruit fly         | SAMN04044078 | SRR2318706                 | 4.4 170 Mb 26 35 1.26 | 10.24 | 127            | 0 2 19 No                               |
| Ectophasia simulans           | Rove beetle       | SAMN0583357  | SRR4301374                 | 22.7 170 1.437 | 42.92 | 127            | 1 3 17 No                               |
| Gerris buenoi 1               | Water striders    | SAMN02439785 | SRR1197265                 | 27.4 990 Mb (GCA_00101 0745.2) 28 69 1.538 | 13.14 | 127            | 0 3 18 No                               |
| Gerris buenoi 2               | Water striders    | SAMN02439786 | SRR1197267                 | 27.8 990 Mb (GCA_00101 0745.2) 28 39 1.537 | 13.14 | 127            | 0 3 18 No                               |
| Homalodisca vitripennis 1     | Glassy-winged     | SAMN02289956 | SRR941995 SRR941996 SRR941997 | 107.7 145 Gb (GCA_0 0069 6855.2) 74 260 1.675 | 14.19 | 121            | 9 3 15 Maybe: BUSCO duplications Lab reared Florida-strain female |
| Macconellicoccus hirsutus     | Mealybug          | SAMEA3699093 | ERR1189167                 | 9.8 160 Mb (GCA_0 0006 4465.1) 61 80 1.415 | 27.94 | 130            | 0 2 16 No                               |
| Opophthora brumata            | Winter moth       | SAMN03121611 | SRR1618545 SRR1618581 SRR1618582 | 22.2 640 Mb (GCA_0 0126 6575.1) 35 28 1.35 | 33.34 | 129            | 0 1 18 No                               |
| Pararge aegeria               | Speckled wood     | SAMN02688782 | SRR1190479                 | 9.8 138 1.282 | 83.56 | 129            | 0 2 17 No                               |
| Pediaspis aceris 1            | Gall wasp         | SAMEA3925672 | ERR1355090                 | 3.8 12 1.188 | 8.62 | 119            | 0 5 24 No                               |
| Pediaspis aceris 2            | Gall wasp         | SAMEA3925673 | ERR1355091                 | 3.8 12 1.174 | 8.05 | 114            | 0 9 25 No                               |
| Polygonia c-album             | Comma butterfly   | SAMN02688783 | SRR1190476                 | 6 61 1.463 | 22.09 | 129            | 0 1 18 No                               |
| Pseudonymyces sp. PSW-54      | Ant               | SAMN03275520 | SRR1742977                 | 36.9 280 Mb (GCA_0 0208 6095.1) 130 195 1.245 | 15.91 | 127            | 0 3 18 No                               |

* Diplolepis spinosa: Gall wasp with estimated ploidy data.
* Drosophila melanogaster: Fruit fly with assembly size and nucleosome size data.
* Drosophila simulans: Fruit fly with different populations and sexes.
* Ectophasia simulans: Rove beetle with assembly size data.
* Gerris buenoi: Water striders with assembly size data.
* Homalodisca vitripennis: Glassy-winged sharpshooter with assembly size data.
* Macconellicoccus hirsutus: Mealybug with assembly size data.
* Opophthora brumata: Winter moth with assembly size data.
* Pararge aegeria: Speckled wood butterfly with assembly size data.
* Pediaspis aceris: Gall wasp with assembly size data.
* Polygonia c-album: Comma butterfly with assembly size data.
* Pseudonymyces sp. PSW-54: Ant with assembly size data.

For Drosophila simulans, there are two entries indicating natural populations and laboratory strains. For Drosophila simulans, there are entries with estimated ploidy and assembly data, including a note indicating that the genome size is estimated as ~150 Mb.
| Organism                  | Genus                  | Genus Information                                                                 |
|--------------------------|------------------------|-----------------------------------------------------------------------------------|
| Rhagoletis pomonella     | Apple maggot fly       | SAMN05388941 SRR3900841 SRR3901027                                              |
| Trichogramma pretiosum   | Wasp                   | SAMN02439301 SRR3900841 SRR3901027                                              |
| Acromyrmex echinatior    | Ant                    | SAMEA762107 ERR034187 ERR03416                                                  |
| Bembidion lapponicum     | Beetle                 | SAMEN04276907 SRR2939026                                                         |
| Biorhiza pallida 3       | Wasp                   | SAMEA2053314 ERR233313                                                          |
| Callistobrachus chinensis| Bean weevil            | SAMN02313283 SRR9409786 SRR952345                                                |
| Ceratina calcarata       | Carpenter bee          | SAMEN04210145 SRR2912519                                                         |
| Cynipini 1               | Oak gall wasp          | SAMEA1965365 ERR233303 ERR233304 ERR233305                                       |
| Cynipini 2               | Oak gall wasp          | SAMEA2053318 ERR233306                                                          |
| Dactylopius coccus       | Domestic cochineal     | SAMN02725055 SRR1251828 SRR1251831 SRR1251832                                  |
| Drosophila melanogaster 3| Fruit fly              | SAMN05417645 SRR3931592                                                         |
| Homalodisca vitripennis 2| Glassy-winged sharpshooter (leafhopper) | SAMN02209957 SRR941998 |
| Isocolus centaurae 1     | Gall wasp              | SAMEA3930555 ERR1359249                                                         |
| Isocolus centaurae 2     | Gall wasp              | SAMEA3930556 ERR1359250                                                         |
| Megacopta cribraria      | Stink bug              | SAMN02313994 SRR1145746                                                         |

**Coverage and Assembly Statistics:**

- **Rhagoletis pomonella**
  - SRR3900841 SRR3901027
  - Coverage: 23.3 C-value: 0.97
  - Length: 25, 270 bp
  - BUSCO: 13.52, 127
  - Annotations: 0, 3
  - Remarks: No, Single adult female fly

- **Trichogramma pretiosum**
  - SRR1191749 SRR1191750 SRR1191751 SRR1191752 SRR1191753
  - Coverage: 68.6 C-value: 0.97
  - Length: 360, 50, 1.097 bp
  - BUSCO: 51.37, 127
  - Annotations: 0, 3
  - Remarks: No

- **Acromyrmex echinatior**
  - SAMEA762107 ERR034187 ERR03416
  - Coverage: 13.2 C-value: 0.75
  - Length: 44, 56, 1.611 bp
  - BUSCO: 4.66, 104
  - Annotations: 0, 8
  - Remarks: Female, 1 male

- **Bembidion lapponicum**
  - Same as **Rhagoletis pomonella**

- **Biorhiza pallida 3**
  - SAMEA2053314 ERR233313
  - Coverage: 4.1 C-value: 0.75
  - Length: 42, 0.645 bp
  - BUSCO: 0.7, 3
  - Annotations: 0, 8
  - Remarks: No

- **Callistobrachus chinensis**
  - SAMN02313283 SRR9409786 SRR952345
  - Coverage: 32.1 C-value: 0.75
  - Length: 44, 340, 2.894 bp
  - BUSCO: 4.06, 78
  - Annotations: 6, 12
  - Remarks: Male, head, thorax, feet

- **Ceratina calcarata**
  - Same as **Rhagoletis pomonella**

- **Cynipini 1**
  - SAMEA1965365 ERR233303 ERR233304 ERR233305
  - Coverage: 9.4 C-value: 0.75
  - Length: 20, 1.216 bp
  - BUSCO: 6.57, 96
  - Annotations: 0, 13
  - Remarks: No

- **Cynipini 2**
  - SAMEA2053318 ERR233306
  - Coverage: 7.3 C-value: 1.053
  - Length: 17, 1.182 bp
  - BUSCO: 3.32, 82
  - Annotations: 0, 21
  - Remarks: No

- **Dactylopius coccus**
  - SAMN02725055 SRR1251828 SRR1251831 SRR1251832
  - Coverage: 6.2 C-value: 21.1
  - Length: 290, 2.563 bp
  - BUSCO: 5.95, 47
  - Annotations: 40, 80
  - Remarks: Male, head, thorax, feet

- **Drosophila melanogaster 3**
  - SAMN05417645 SRR3931592
  - Coverage: 3.4 C-value: ~175
  - Length: 19, 6.5 bp
  - BUSCO: 6.14, 102
  - Annotations: 0, 15
  - Remarks: Adult male whole body, wild caught from Africa

- **Homalodisca vitripennis 2**
  - SAMEA2209957 SRR941998
  - Coverage: 39.5 C-value: 1.45
  - Length: 27, 1.803 bp
  - BUSCO: 17.02, 109
  - Annotations: 23, 3
  - Remarks: Lab-reared Florida-strain male

- **Isocolus centaurae 1**
  - SAMEA3930555 ERR1359249
  - Coverage: 3.4 C-value: 0.98
  - Length: 7.9, 0.986 bp
  - BUSCO: 2.19, 63
  - Annotations: 0, 16
  - Remarks: No

- **Isocolus centaurae 2**
  - SAMEA3930556 ERR1359250
  - Coverage: 3.4 C-value: 0.98
  - Length: 7.7, 0.965 bp
  - BUSCO: 2.31, 63
  - Annotations: 0, 14
  - Remarks: No

- **Megacopta cribraria**
  - SAMN02313994 SRR1145746
  - Coverage: 5.7 C-value: 1.053
  - Length: 31, 2.097 bp
  - BUSCO: 1.74, 73
  - Annotations: 4, 13
  - Remarks: Male, head, thorax, feet

**Comments:**

- **Rhagoletis pomonella** and **Trichogramma pretiosum** share high BUSCO values and similar assembly sizes.
- **Drosophila melanogaster** shows high BUSCO and assembly coverage, indicating a high-quality assembly.
- **Homalodisca vitripennis** and **Isocolus centaurae** show lower BUSCO values, suggesting potential assembly and duplication issues.
- **Megacopta cribraria** demonstrates bimodal coverage distribution and possible assembly size variations.
| Species                  | Family        | GenBank Accession Numbers | Length (bp) | Coverage (×) | N50 (bp) | GC Content (%) | Chimeric (%) | BUSCO Duplicates | Notes                                                      |
|-------------------------|---------------|---------------------------|-------------|--------------|----------|----------------|--------------|------------------|------------------------------------------------------------|
| Mycopsylla fici 1       | Fig psyllid   | SAMN04226368              | 0.9         | 9.9          | 1.171    | 2.32           | 76           | 0                | 6                | 66                | No                                                         |
| Mycopsylla fici 2       | Fig psyllid   | SAMN04226369              | 0.8         | 12           | 1.238    | 2.68           | 77           | 0                | 14               | 57                | No                                                         |
| Mycopsylla proxima      | Psyllid       | SAMN04226370              | 1.1         | 5.7          | 0.364    | 0.76           | 0            | 0                | 3                | 145               | No                                                         |
| Rhagoletis zephyria     | Tephritid fly | SAMN04977950              | 132.9       | 1.3 Gb       | 1.881    | 11.19          | 80           | 51               | 2                | 15                | Yes: bimodal coverage distribution; assembly size; BUSCO duplications | Single adult female fly |