bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title

Deep data mining reveals variable abundance and distribution of microbial reproductive manipulators within and among diverse host species

Authors

Paloma Medina[1]*, Shelbi L Russell[2], and Russell Corbett-Detig[1]*

Affiliations

[1] Genomics Institute and Department of Biomolecular Engineering UC Santa Cruz, Santa Cruz, CA 95064

[2] Department of Molecular, Cellular and Developmental Biology, UC Santa Cruz, Santa Cruz, CA 95064

* Correspondence To: pamedina@ucsc.edu, rucorbet@ucsc.edu

Abstract

Bacterial symbionts that manipulate the reproduction of their hosts to increase their successful transmission are important factors in invertebrate ecology and evolution. In light of their use as a biological control agent, studying the genomic and phenotypic diversity of reproductive manipulators can improve efforts to control infectious diseases and contribute to our understanding of host-symbiont evolution. Despite the vast genomic and phenotypic diversity of reproductive manipulators, only a handful of *Wolbachia* strains are used as biological control agents because little is known about the broad scale infection frequencies of these bacteria in nature. Here we develop a data mining approach to quantify the number of arthropod and nematode host species available on the Sequence Read Archive (SRA) that are infected with *Wolbachia* and other reproductive manipulators such as *Rickettsia* and *Spiroplasma.* Across the entire database, we found reproductive manipulators infected 1733 arthropod and 103 nematode samples, representing 121 and 10 species, respectively. We estimated that *Wolbachia* infects approximately 24% of all arthropod species and 20% of all nematode species. In contrast, we estimated other reproductive manipulators infect 0-8% of arthropod and nematode species. We show that relative *Wolbachia* density within hosts, titer, is significantly lower than the titer of the other reproductive manipulators. Considering the fitness costs of high titers, low titer may contribute to enabling *Wolbachia's* high prevalence across hosts species and mitigate impacts on host biology compared with other reproductive manipulator taxa. Our study demonstrates that data mining is a powerful tool for understanding host-symbiont co-evolution and opens an array of previously inaccessible questions for further analysis.

Introduction

Bacterial symbionts of eukaryotic hosts exhibit an impressive array of phenotypes that interact with host biology. Included among these symbionts are bacteria that alter host reproduction in order to increase their likelihood of transmission to the next host generation [1–4], a strategy termed reproductive manipulation. Depending on the nature of host reproduction, reproductive manipulators are generally transmitted vertically by associating with either the oocyte or developing embryos [5]. Manipulative phenotypes range from strategies that prevent the survival of uninfected offspring, such as cytoplasmic incompatibility, to strategies such as feminization, male killing, and parthenogenesis, that actually change the sex ratio to favor females for overall increased infection rates [1–4]. These drastic manipulations to normal host biology mean that reproductive manipulators can induce reproductive isolation, drive changes in sexuality, and alter reproductive ecology in their hosts [6–8].

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

While reproductive manipulators undergo vertical transmission to reach the next host generation by definition, many species and strains also exhibit low rates of horizontal transmission between contemporary host species. This combination of transmission strategies coupled with reproductive manipulation serves to quickly increase the prevalence of reproductive manipulators in a host population and to spread rapidly among host species. Additionally, some reproductive manipulators are mutualistic, helping to increase host fecundity, and further increasing infection rates [9-12]. For example, the strain of *Wolbachia* that infects bed bugs supplements B vitamins lacking in the host's diet [9] and some strains of *Wolbachia* protect hosts from viral or parasite infection [10, 11].

Given the impressive abilities of these bacteria to colonize new hosts and impact host fertility and development for their own reproduction, much effort has been spent attempting to estimate the infection frequencies of reproductive manipulators. Due to its wide distribution and presence in model organisms such as *Drosophila, Wolbachia* is one of the most well studied endosymbionts in general, and is the best studied reproductive manipulator in particular. Estimates of the frequency of *Wolbachia* infections amongst arthropods range from 11% [13] to 76% [14]. The large variation in estimates is likely an effect of sampling bias and variation in assay sensitivity. For example, PCR amplification of the *Wolbachia* surface protein gene (*wsp*) is commonly used to estimate the infection frequency and species distribution of this reproductive manipulator [15–17]. However, *wsp* is not present in all *Wolbachia* strains, and thus, the use of these primer sets biases the data and underestimates infection rates [18]. Other reproductive manipulators in the *Cardinium, Arsenophonus, Rickettsia* and *Spiroplasma* clades are reported to occur in 4% to 7% of all species [19]. However, these estimates are less certain than *Wolbachia* because less data is available for these clades, and similar detection biases might impact frequency estimates. Given that the probability of sampling an infected individual is positively correlated to the prevalence of the endosymbiont and the number of individuals sampled from the host population, undersampling also imposes a barrier to confident detection of low frequency infections.

Reproductive manipulators also experience a range of fitness tradeoffs during host growth and development. A symbiont must be present at sufficiently high frequencies within a host to promote successful transmission to subsequent host generations. However, exceedingly high abundance of symbiont cells relative to host cells, termed titer, may impose a significant fitness cost for the host and their endosymbionts [20, 21]. Reproductive manipulators may be more virulent to their hosts when they achieve high titers, and natural selection on symbiont proliferation and host regulation should favor the evolution of intermediate endosymbiont frequencies [22, 23]. This tradeoff between reliable transmission and fitness costs to hosts is therefore an essential component of understanding reproductive manipulator co-evolution. Nonetheless, little is known about the relative abundances of reproductive manipulators within host individuals in large part due to the challenges of collecting these data in high throughput ways.

Characterizing the prevalence and distribution of reproductive manipulators could be especially valuable to biomedical researchers using reproductive manipulators to control the spread of human pathogenic viruses such as Zika, Dengue and Chikungunya [24–26]. Currently, only the wMel strain of *Wolbachia* is being used as a biological control agent [26, 27]. However, the large genetic and phenotypic diversity of reproductive manipulators could suggest different species or strains of bacteria to combat the spread of arboviruses. Additionally, a strategy that opportunizes on cytoplasmic incompatibility, termed the "sterile male technique", can be used to control mosquito population sizes [28, 29]. As this strategy requires low infection frequencies in females to work [30] and different strains may have compatible rescue abilities [31], it is necessary to

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

understand the diversity and distribution of these bacteria in nature. Cataloguing the distribution and titer of reproductive manipulators builds a foundation to explore reproductive manipulators as powerful and versatile biological control agents.

Whole genome sequencing and bioinformatic approaches offer appealing alternatives to conventional PCR-based survey methods to estimate reproductive manipulator infection rates. These methods are not biased by primer selection, are less sensitive to false positives due to contamination, and enable testing of large numbers of samples. Similarly, genome-sequencing is a potentially powerful tool for interrogating endosymbiont titer within host individuals. With Illumina shotgun sequencing, when the genome of a potential host individual is sequenced, the host's endosymbionts are also sequenced. This makes the publicly-available databases a treasure trove for sampling reproductive manipulators with bioinformatic approaches. For example, the NCBI Sequence Read Archive (SRA) [32] contains ~62,000 sequencing runs for samples tagged as Arthropoda alone. Searching, or mining, genomic sequencing data has been shown to be a cost effective and powerful strategy to detect *Wolbachia* infections [13, 33]. However prior studies did not include other reproductive manipulators [13, 33] or were focused on a single host species [33]. Indeed, one of the most compelling arguments for using non-targeted, publicly available data is to avoid ascertainment bias towards selecting species already known to harbor reproductive manipulator infections.

Here, we develop and apply a computational pipeline to determine the infection status of host samples downloaded from the SRA database and estimate the titer of endosymbionts within positively infected host individuals. Using this approach, we comprehensively survey for reproductive manipulators across diverse host species and estimate their global infection frequencies. We classified 1733 arthropod and 103 nematode samples as infected with a reproductive manipulator, including at least 54 previously unknown infections. Additionally, we show that endosymbiont titer varies systematically across bacterial clades and may partly explain the disparate global distributions of reproductive manipulators.

Results and Discussion

Arthropod and Nematode Infections in the SRA

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

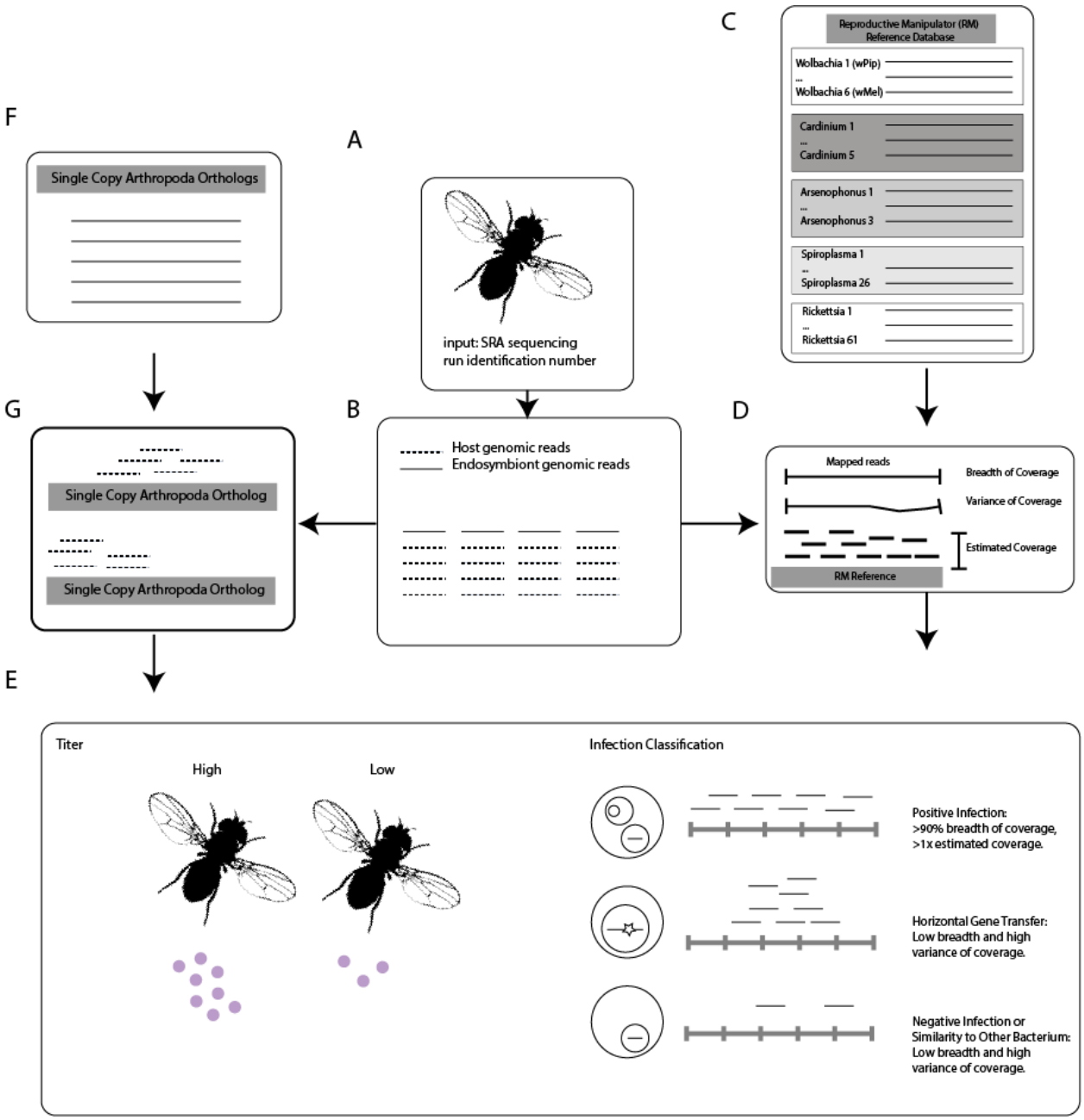


Figure 1. A schematic of the computational pipeline used to determine the reproductive manipulator infection status and endosymbiont titer of a sequencing run (see also Supplementary Methods S1 and S7). The pipeline **(A)** takes in a sample's unique identification number, then **(B)** downloads two million endogenous and exogenous reads. Then, **(C)** reads are aligned to *Wolbachia, Arsenophonus, Spiroplasma, Cardinium,* and *Rickettsia* reference genomes, and **(D)** summary statistics for the sample aligned to each reference are computed. If a sample had between 0.1 and 0.9 breadth of coverage, the full dataset was downloaded and the workflow repeated to prevent false negative calls. **(F and G)** To estimate host coverage without requiring a reference genome, reads are also aligned to a set of 1066 single copy ancestral orthologs obtained from ORTHODB v9. Then, **(E)** we apply coverage breadth and depth cutoffs to classify infection status as positive, or negative. We compare the coverage of host reads to single copy orthologs to the coverage of endosymbiont reads to reproductive manipulator reference genome to estimate endosymbiont load.

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We developed a powerful bioinformatic pipeline to identify reproductive manipulator infections within sequencing datasets. Briefly, our approach compares sequencing reads from a given sample to a set of reference genomes from reproductive manipulator species to determine if a given host is positively infected (Figure 1, Methods). We extensively characterized the sensitivity and accuracy of our bioinformatic pipeline using previously known *Wolbachia* infection statuses of the Drosophila Genetic Reference Panel (DGRP [34, 35]) and other samples known to harbor genetically divergent *Wolbachia* [36]. Using the confident infection status lines in the DGRP as our validation set, our method is completely concordant with both PCR and previous bioinformatic methods. Additionally, our approach remains accurate even for references that exhibit 5-15% pairwise sequence divergence (Tables S5 and S6, Figures S5 and S6). These results indicate that our pipeline is an efficient and accurate method to determine reproductive manipulator infections in the majority of host samples (Supplementary Method S5).

Using our bioinformatic classification pipeline (see Methods, Figure 1), we tested nearly all arthropod and nematode samples on the SRA database that had been whole genome shotgun sequenced (as of November 8, 2017) and found 121 arthropod species and 10 nematode species had samples infected with a reproductive manipulator out of 1,259 and 128 species tested, respectively (Figure S1 and S2, Table S1 and S3, File S1 and S2). We identified 54 arthropod species with previously unreported reproductive manipulator infections (Table S2, Supplementary Method S2). *Wolbachia* was the most frequent reproductive manipulator in arthropod samples, and was the sole reproductive manipulator found in nematodes, as expected based on previous work [19, 37]. Almost all nematode species infected with *Wolbachia* are filarial worms (Table S3), a result supported by previous studies showing that filarial nematodes and *Wolbachia* are in an obligatory, mutualistic relationship [38–40]. There was one non-filarial nematode species infected with *Wolbachia*: *Pratylenchus penetrans*, a plant-parasitic nematode (order Tylenchida). This species has been shown previously to be infected with *Wolbachia* [37, 41]. These results in addition to our validation experiments indicate our approach can accurately detect endosymbiont infections.

We also found numerous species infected with other clades of reproductive manipulators (Table S1). Specifically, *Rickettsia*, *Spiroplasma*, and *Arsenophonus* infect 28, 18, and four arthropod species in our dataset, respectively. In nematodes, we found only a single sample infected with *Cardinium*. The substantially lower infection rates for other reproductive manipulators than for *Wolbachia* are consistent with prior work [42–44] (Table S3, Figure S2).

We also found high rates of co-infections of different reproductive manipulators among arthropod hosts species and individuals. Species co-infections occurred in 15 of the 140 arthropod species that harbored any infection ($p < 0.001$ permutation test, Supplementary Method S6, Figure S3, and Table S4). We also found an excess of individual samples infected by more than one reproductive manipulator, a moderately significant excess relative to permutations (17 individuals, $p < 0.025$ permutation test, Figure S4). These results may indicate that a subset of hosts, both at the species level and within single host maternal lineages, are more likely to acquire endosymbionts than others or that they are more permissive to prolonged infections. This is supported by the observation that some host species harbor genetic variation that influences the abundance of endosymbionts within individuals [22, 45]. Alternatively, if these bacteria influence the host environment to be more permissive to microbial endosymbionts, this might facilitate the build-up of reproductive manipulators within a single species or host maternal lineage (*i.e.*, a type of niche-construction).

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Estimates of global infection frequencies.
Our study aims to estimate the infection rate of reproductive manipulators in arthropod and nematode species overall, however, different taxa have been sampled to varying degrees and infection frequencies vary (Table S1) so the number cannot be directly calculated from infection counts of individuals. To evaluate and correct for this ascertainment bias, we use a beta-binomial distribution to estimate the total proportion of species infected with a reproductive manipulator bacterial species as has been done previously [46] (Supplemental Method S3, Figure S7). Using this approach, we estimated that 24% (95% CI 11-38%), and 20% (95% CI 3-52%), of arthropod and nematode species, respectively, are infected with *Wolbachia* (Table 1). We note that these values are consistent with expectations from previous work [14, 47–53] and emphasize that because the SRA has been populated with samples mostly without considering *Wolbachia* infection status, this approach should provide a relatively unbiased estimate of global infection frequency.

Among the other reproductive manipulators, we estimated *Arsenophonus, Rickettsia, and Spiroplasma* infected 4%, 5%, and 9% of arthropod species, respectively (Table 1). We were not able to estimate global infection rates of *Cardinium* because of the extremely low rate of positive infection in our dataset. Owing to the fact that we do not have positive and negative controls readily available for each of these other reproductive manipulator clades, it is difficult to completely rule out infections that failed to map to the known reference(s) for each group and therefore induce a higher rate of false negatives. However, our results from mapping *Wolbachia* reads to extremely diverse reference genomes (e.g., 5-15% divergence) suggests that the rate of false negatives is low, provided the divergence within these other bacterial groups does not exceed our tested values and, as we note above, our raw frequency estimates are in line with previous work based on other methods.

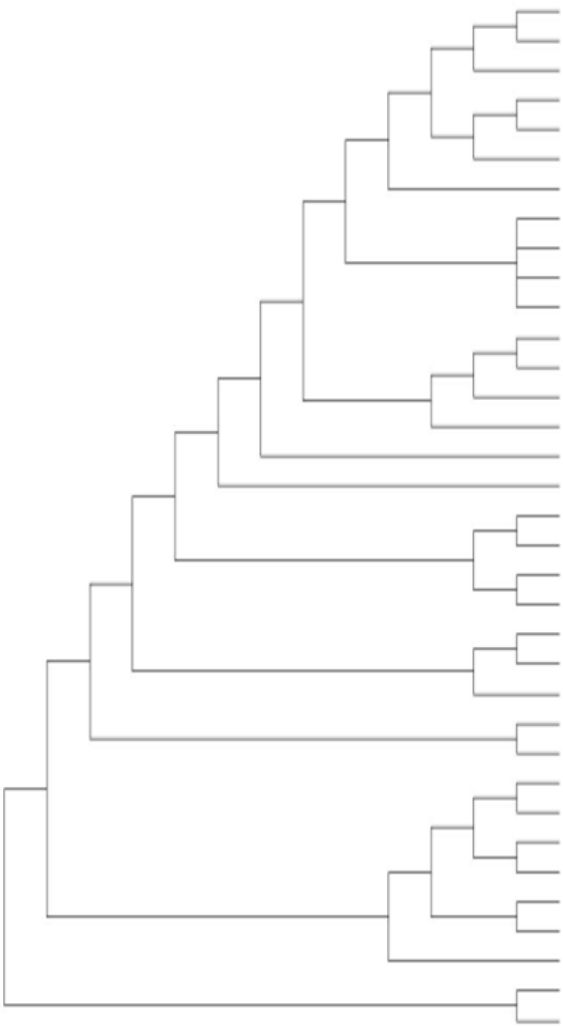
Table 1. Estimated infection frequencies and confidence intervals from our data for *Wolbachia, Spiroplasma,* and *Arsenophonus* infecting arthropods and nematodes. All species in dataset were downsampled to maximum 100 individuals. We used a minimum threshold of 1 in 1000 of infected individuals within a species to classify a species as positively infected (Supplementary Method S3).

Host Phylum	Reproductive Manipulator	2.5% CI	97.5% CI	Msedian	Mean
	Arsenophonus	0.0002	0.2089	0.0170	0.0368
	Rickettsiales	0.0145	0.1418	0.0462	0.0542
	Spiroplasma	0.0200	0.2846	0.0688	0.0855
Arthropods	*Wolbachia*	0.1191	0.3847	0.2365	0.2407
Nematodes	*Wolbachia*	0.0266	0.5225	0.1779	0.2035

Phylogenetic Distribution of Reproductive Manipulator Infections
The phylogenetic distribution of *Wolbachia* spans 11 arthropod orders, out of the 35 tested (Figure 2). Across all of the arthropod species that we studied here, the orders with the greatest number of species sampled are Hymenoptera, Diptera, Lepidoptera, Coleoptera, and Hemiptera. These orders had 470, 164, 104, 96, and 55 species sampled respectively. Positively infected samples are dispersed widely across the tree, including hosts as distant as Coleoptera and Araneae. Horizontal transmission among species may contribute towards

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

explaining *Wolbachia's* large phylogenetic distribution despite only modest sequence level divergence.



Arthropoda Order	Species Sampled	Species Wolbachia Positive	Species Arsenophonus Positive	Species Rickettsia Positive	Species Spiroplasma Positive
Strepsiptera	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Coleoptera	96	7 (0.073)	0 (0.000)	13 (0.135)	0 (0.000)
Neuroptera	1	0 (0.000)	0 (0.000)	1 (1.000)	0 (0.000)
Trichoptera	5	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Lepidoptera	104	6 (0.058)	0 (0.000)	1 (0.010)	5 (0.048)
Diptera	164	27 (0.165)	0 (0.000)	2 (0.012)	4 (0.024)
Hymenoptera	470	43 (0.091)	2 (0.004)	3 (0.006)	4 (0.009)
Thysanoptera	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Psocoptera	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Phthiraptera	62	0 (0.000)	0 (0.000)	1 (0.016)	0 (0.000)
Hemiptera	35	11 (0.314)	2 (0.057)	3 (0.086)	0 (0.000)
Blattodea	12	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Phasmatodea	3	1 (0.333)	0 (0.000)	0 (0.000)	0 (0.000)
Orthoptera	11	1 (0.091)	0 (0.000)	0 (0.000)	0 (0.000)
Plecoptera	4	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Ephemeroptera	3	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Odonata	2	1 (0.500)	0 (0.000)	0 (0.000)	0 (0.000)
Harpacticoida	2	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Calanoida	2	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Notostraca	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Diplostraca	4	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Amphipoda	4	1 (0.250)	0 (0.000)	0 (0.000)	0 (0.000)
Isopoda	4	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Decapoda	34	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Spirobolida	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Geophilomorpha	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Astigmata	3	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Trombidiformes	2	1 (0.500)	0 (0.000)	1 (0.500)	0 (0.000)
Ixodida	4	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Mesostigmata	3	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Araneae	8	1 (0.125)	0 (0.000)	0 (0.000)	0 (0.000)
Scorpiones	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Xiphosura	3	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Poales	1	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)
Asterales	2	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)

Figure 2. Phylogeny of Arthropoda orders tested and number of reproductive manipulator-positive species within each order. The frequency of reproductive manipulator-positive species listed in parentheses. We used the Tree of Life taxonomic and phylogenetic package [54], rotl [55], to group host species by their orders.

Wolbachia infection frequencies vary substantially across insect orders. Hymenoptera, Diptera, and Coleoptera have the highest global *Wolbachia* infection frequency, each around 37% of species infected (Table 2). Conversely, We observed a relatively low global *Wolbachia* infection frequency estimates for Hemiptera (22%) and Lepidoptera (1.5%, Table 2). Lepidoptera's infection frequency, in particular, is significantly lower than the frequencies found in other orders of insects (p < 6e-3, permutation test). Low frequencies might be the result of high fitness cost to infected hosts or from *Wolbachia's* low transmission fidelity in these groups. Notably,

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wolbachia infections feminize and kill males in Lepidopteran hosts. Feminization imposes a fitness cost on populations harboring these endosymbionts because fewer males are available for mating and because males tend to prefer genetic females over feminized males, which have lower mating rates and receive less sperm [56]. Lower infection frequencies of sex-ratio-distorting phenotypes might suggest feminization and male killing have a higher fitness cost, or that host populations can more easily subvert sex-ratio-distorting phenotypes compared to other reproductive phenotypes like cytoplasmic incompatibility and parthenogenesis. Sex-ratio-distorting phenotypes might therefore play a role in limiting infection polymorphism in host populations.

Table 2. *Wolbachia* global infection frequencies and confidence intervals generated for arthropod orders. All species in dataset were downsampled to maximum 100 individuals. Confidence intervals were generated using 1000 bootstrap replicates fitting a beta-binomial model to species infection frequency data among orders. A minimum infection frequency of 0.001 was used to classify a species as positively infected (Supplementary Method S3).

Order	2.50%	97.50%	Median	Mean
Coleoptera	0.0173	0.8685	0.3956	0.3981
Diptera	0.0956	0.7074	0.3818	0.3868
Hymenoptera	0.1227	0.7362	0.3607	0.3832
Hemiptera	0.0357	0.7315	0.1578	0.2215
Lepidoptera	0.0016	0.0476	0.0114	0.0146

Suitability and Properties of our Approach for Classifying Infection Status

Our reference-based method is widely applicable and can be used to detect other symbionts, such as medically relevant pathogens, for which only distantly related references are available. Our approach, which uses whole reference genomes and permits alignment mismatches, stands in contrast to other approaches which only use a few marker loci and require exact matches between reference and sample sequences (e.g., [13]). Thus, we expect that our method is simultaneously more resistant to false positives (by requiring a minimum genome alignment coverage breadth) and false negatives (by being robust to moderate levels of bacterial genetic diversity). Also, our pipeline could be modified to detect more than just bacterial infections (e.g., viral components), and can incorporate any number of reference genomes.

Endosymbiont Load

Replication control is important for vertically transmitted endosymbionts because the fitness of the symbiont is dependent on the fitness of the host. Reproductive manipulator infections must remain sufficiently high to ensure vertical transovarial transmission while being low enough to reduce pathogenic cost to a host [57]. Thus replication control is important for reproductive manipulators to be maintained in host populations. However, the extent to which reproductive manipulator titers vary between orders of arthropods and among reproductive manipulator clades is largely unknown and might be an important component of the fitness costs reproductive manipulator species impose on their arthropod hosts.

Wolbachia* load comparison between *D. melanogaster and D. simulans.

To evaluate the accuracy of our titer estimation method (Supplementary Method S7), we compared the density of *Wolbachia* within *D. melanogaster* and *D. simulans*, which are two of the most commonly studied species. Titer of *Wolbachia* influences its degree of virulence on its host

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[58]. We hypothesized that titers of *Wolbachia* infecting *D. melanogaster* would be significantly different than the titer of strains infecting *D. simulans* for three reasons. First, the strength of reproductive manipulation in *D. simulans* is stronger than *D. melanogaster* [59, 60] the wRi strain commonly found in *D. simulans* has been reported to exhibit higher titers during embryogenesis than wMel does in *D. melanogaster* [61]. Third, wMel transinfected into *D. simulans* exhibited significantly higher titers than in its native host, *D. melanogaster* [20].

We found that the strains inhabiting *D. simulans* exhibit significantly higher titer relative to those in *D. melanogaster* (p < 1e-05, one-tailed Mann Whitney U Test, Figure 3). In fact, for some extreme samples, titer in *D. simulans* shows an astounding >30:1 bacterial to host haploid genome complement ratio (Supplementary Method S7). Accounting for relative genome sizes, this indicates that *Wolbachia* contributes approximately 30% as much DNA as the host does to the sequence data for that individual. Furthermore, we compared the amount of fluorescence due to *Wolbachia* DNA staining between stage9/10a developing oocyte cysts from *D. melanogaster* flies infected with wMel and *D. simulans* flies infected with wRi, and found an average 2.1x more signal in *D. simulans* than *D. melanogaster* (n=15 and 19, respectively, Wilcoxon Rank-Sum p < 5.12e-05; Figure 3, Supplementary Method S4). Thus, the bioinformatically predicted titer differences between *Wolbachia* strains in *D. melanogaster* and *D. simulans* are also reflected in the female germline. Importantly, this suggests that our approach can yield information about *Wolbachia* abundance in tissues that are most relevant to understanding *Wolbachia* transmission.

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

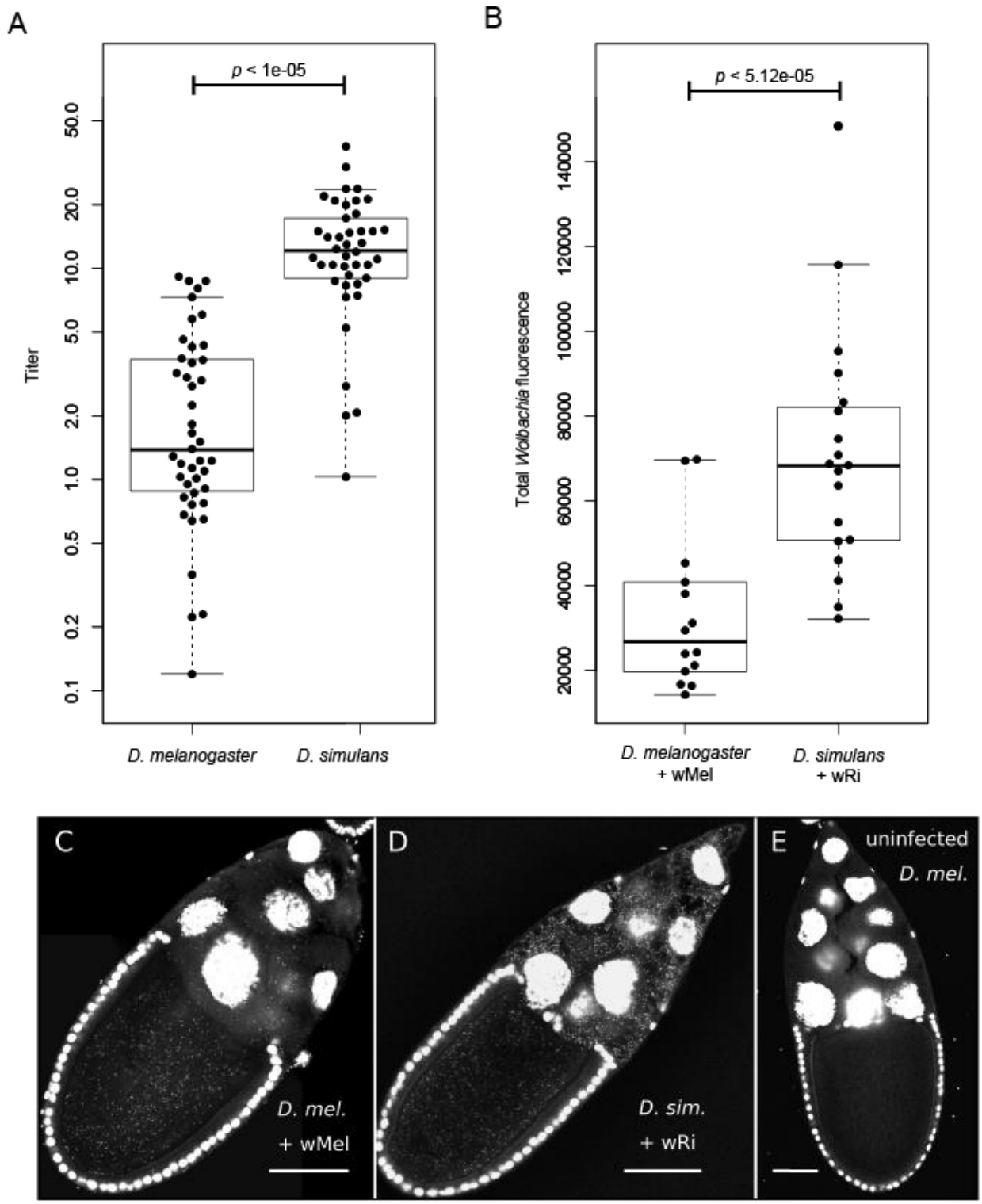


Figure 3. *Wolbachia* titer among *D. melanogaster* and *D. simulans* computed from our **(A)** *in silico* titer estimate approach (y-axis log10 scaled) and **(B-E)** *in vivo* fluorescence assay in stage 10a oocyte cysts. **(B)** Quantification of *Wolbachia* fluorescence, showing a significant difference in titer between host species (Wilcoxon rank sum test). **(C)** *D. melanogaster* and **(D)** *D.*

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

simulans imaged propidium iodide DNA staining and confocal microscopy. **(E)** Uninfected *D. melanogaster* oocyte shows no evidence of *Wolbachia* staining. Scale bars = 50um.

Comparative study of titer among symbiont species

While endosymbiont titer does not necessarily predict the type or intensity of reproductive manipulation phenotypes in insect hosts [58], it does generally predict the virulence or cost of infection to a host [21, 22]. High titers are necessary to ensure an adequate number of bacteria make it to the germline for reliable vertical transmission. However, too high of titer may negatively impact host fitness. Since endosymbiotic bacteria like *Wolbachia* are maternally inherited, strong fitness costs to the host would also impact the fitness of *Wolbachia*. Therefore, theory predicts symbiont titer will evolve towards a “goldilocks” zone, minimizing fitness costs to host and maximizing vertical transmission of symbiont bacteria [62]. The mechanism for this could either be via selection on endosymbiont growth or activity itself [63, 64] or selection on host sanctions and regulatory mechanisms employed on the endosymbionts [22, 23, 64, 65].

We hypothesize that *Wolbachia* titer in arthropod hosts would be significantly different compared to other reproductive manipulator bacterial titers because *Wolbachia* is found at substantially higher frequencies across arthropod species and many studies have reported that *Wolbachia* generally elicits smaller fitness effects on its hosts than do *Rickettsia, Arsenophonus,* and *Spiroplasma* [66–71]. In comparing titer across samples infected with each bacterial clade, we found that *Wolbachia* is present at a significantly lower titer in its hosts than *Rickettsia, Spiroplasma,* and *Arsenophonus* (p < 1e-4 for all comparisons, Mann Whitney U) (Figure 4). There was no significant difference between the titers of any other pair of reproductive manipulators. Wolbachia’s relatively low titer may cause differential fitness for this group relative to the other reproductive manipulators. This is consistent with the reported positive relationship between titer and virulence [21, 22]. Our results therefore suggest that *Wolbachia*’s relatively low titer reduces fitness costs to its hosts, thereby contributing to *Wolbachia*’s success and its high prevalence compared to the other reproductive manipulators tested.

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

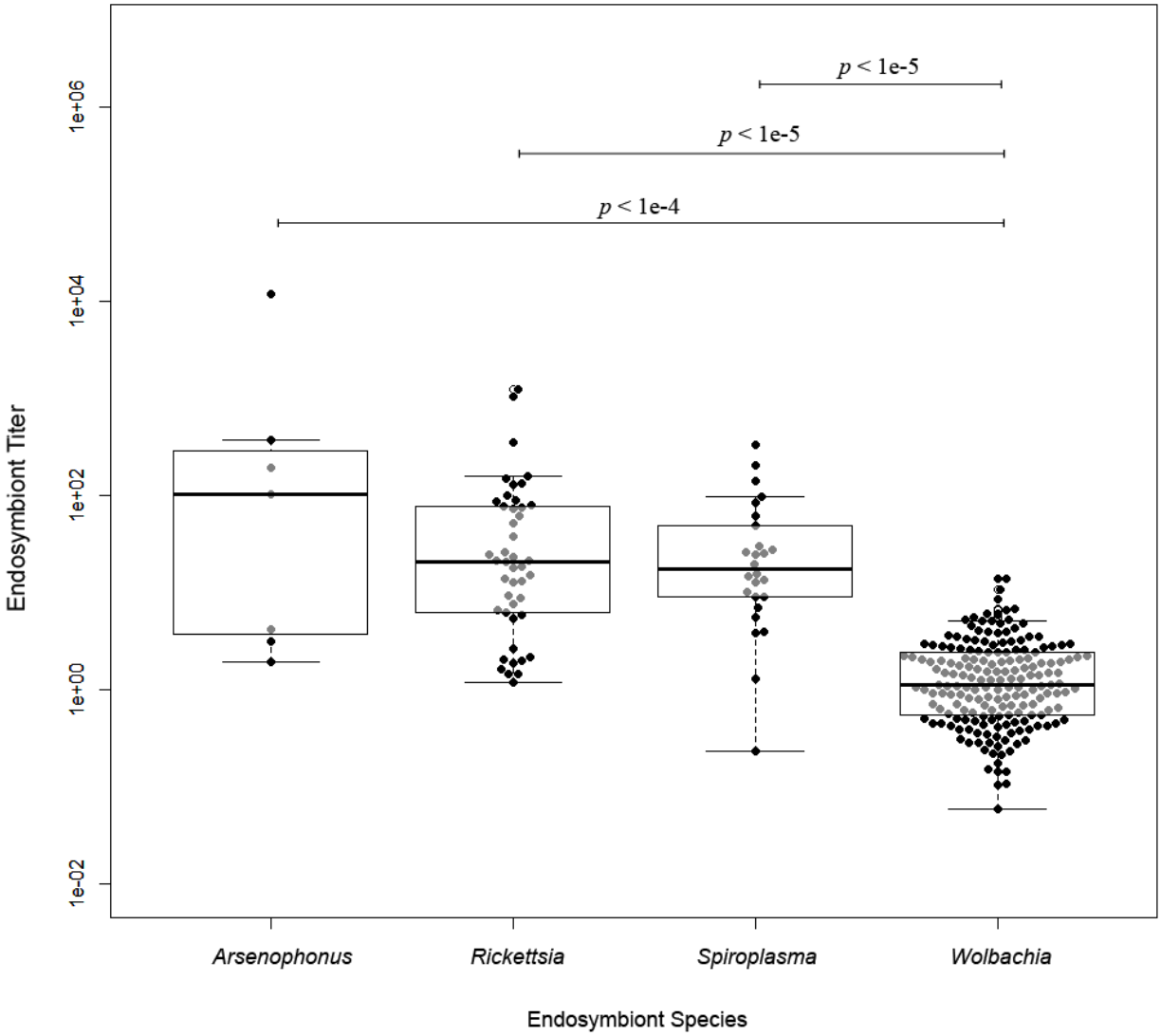


Figure 4. Titer up to three individual per positively infected host species for *Wolbachia*, *Rickettsia, Arsenophonus, Spiroplasma*. Titer values were calculated by comparing endosymbiont genome coverage with host nuclear coverage, and were plotted as swarm plots, overlaid with boxplots. The axis is log10 scaled.

Associations between *Drosophila melanogaster* genotype and *Wolbachia* titer

Considering the strong influence *Wolbachia* strains can have on their hosts' reproductive outcomes, it is likely that the host genome experiences strong selective pressures to respond to or control the bacteria. Indeed, previous work by [22] identified a gene in the wasp host, *Nasonia vitripennis,* that effectively controls *Wolbachia* titer and has been under recent positive selection, likely since the wVitA strain horizontally transferred into the species. To address this question using our public data-sourced method, we used the DGRP to perform a genome-wide association study (GWAS) of titer in *D. melanogaster.*

Although the present analysis is exploratory, there are several noteworthy results. We identified 16 candidate single nucleotide polymorphisms (SNPs) associated with *Wolbachia* titer (Table S9).

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

One of the most strongly associated SNPs is found in a gene associated with the endoplasmic reticulum membrane (*XBP1*) [72]. Consistent with this functional prediction, a recent study on a wMel-infected *D. melanogaster* cell line found that *Wolbachia* resides within membrane derived from the ER, and resides near and within this organelle [73]. Modifying this membrane might therefore enable the host to impact *Wolbachia* titer. Additionally, another strongly associated SNPs is found in a gene associated with actin binding and microtubule transport (CG43901). Recent work has shown that Wolbachia uses host actin for localization in host tissues [74, 75] and modifications to actin-binding proteins can clear *Wolbachia* infections in host individuals [76]. Finally, a SNP in CG17048 is also strongly associated with *Wolbachia* titer in the DGRP. This gene's role in protein ubiquitination is consistent with previous results from a genome-wide RNAi screen that found this process contributes disproportionately to modifications of *Wolbachia* titer in cell culture [77]. These are therefore appealing candidate genes for evaluating the potential of natural host variation to control endosymbiont infections.

Functional work will be necessary to validate our specific predictions, nonetheless these titer-associated genetic polymorphisms suggest that the host genome is capable of evolving to control *Wolbachia* infections. Although the present association work is focused on a well-characterized genetic mapping panel in *D. melanogaster*, our results illustrate more generally the potential impacts of high throughput data mining for identifying both reproductive manipulators and candidate host genetic factors that control them more broadly.

Conclusion

We present a reference-based approach to detect bacterial infections in short read data and highlight the ways it can be used to generate insights into host-symbiont interactions. Our work is the first to estimate the global distribution of multiple reproductive manipulators across their potential arthropod and nematode hosts using full-genome high throughput methods. Moreover, we show that publically-available short read data can be used to interrogate other biological attributes of host-endosymbiont associations, such as titer. Through these analyses, we found significant correlations among frequencies and taxa, as well as titers and taxa. For both of these measures, *Wolbachia* is the outlier among reproductive manipulators, exhibiting both higher infection frequencies and lower titers than the other bacteria. Our results therefore indicate that *Wolbachia's* privileged role as the most common microbial endosymbiont taxon is far from accidental and instead suggests that a substantial co-evolutionary process has generated its widespread distribution. Our database catalogs some of the vast phenotypic and genetic diversity of reproductive manipulators and contributes a diverse annotation of reproductive manipulator strains that can be used as biological control agents to control the spread of infectious disease.

While our approach relies on datasets gathered for a wide array of purposes and therefore requires a level of approximation, we have shown that accurate and precise predictions can still be obtained using this method. Moreover, as publicly available sequence data continues to accumulate at exceptional rates, this framework will become increasingly powerful relative to gathering purpose built-datasets to assay endosymbiont infection statuses and frequencies. More generally our method and related approaches could be used to detect other microbial symbionts, such as medically relevant pathogens, or even viruses, for which a reference genome sequence is available. Hence, future work will build on this framework of leveraging increasingly vast datasets to conduct direct and precise hypothesis testing of fundamental questions interrogating host and endosymbiont ecology and evolution.

Methods

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Reproductive manipulator reference genome panel

We built our BLAST database using RefSeq genome assemblies for *Arsenophonus, Spiroplasma, Rickettsia,* and *Wolbachia.* We included three *Arsenophonus*, six *Wolbachia,* 27 *Spiroplasma*, five *Cardinium,* and 61 *Rickettsia* genome assemblies (Table S10). These genomes were selected to span the known diversity of these bacterial groups. On average, each genome assembly was about 1.3 Mb. We used these 102 genome assemblies to build a BLAST database using the blastdb command from the NCBI blast package (version 2.7.1).

Arthropod and nematode SRA dataset

We downloaded all Arthropoda sequencing read data from the NCBI Sequence Read Archive (SRA) database [32]. We filtered samples under the Arthropoda and Nematoda taxonomy for those sequenced on Illumina sequencing platforms. We filtered nominally for whole genome shotgun libraries, but for completeness we further removed samples that were marked as "reduced representation", "chipseq", and other terms that preclude a shotgun library approach. In total, we tested 18,540 arthropod and 4,671 nematode samples for reproductive manipulator infections. We consolidated all subspecies into a single species which resulted in a total of 1,259 arthropod and 128 nematode species. SRA metadata for samples we screened can be found in Supplemental Tables S11 and S12.

Determining Infection Status using a BLAST-based approach.

In order to classify a host sample as positively or negatively infected with a reproductive manipulator, we analyzed local alignments between host DNA sequence data and reproductive manipulator reference genomes. We binned each reference genome into 5kb segments and computed the proportion of bins with a significant hit (breadth of coverage). We also computed the variance of BLAST hits across each bin and estimated the coverage of the reproductive manipulator genome (Supplementary Method S1). We determined a sample to be a candidate for a positive infection if it had a 90% breadth of coverage and >1x estimated coverage on a reproductive manipulator reference genome.

Validation of bacterial detection pipeline

To estimate the sensitivity and the specificity of our reproductive manipulator annotation method, we compared our results to a previous extensive survey conducted by [33], which was PCR validated. We determined the reproductive manipulator infection status of 158 individuals from the *Drosophila Genetics Research Panel* (DGRP) using our computational pipeline. These 158*D. melanogaster* samples had matching PCR and WGS determined infection statuses generated by [33] and had their sequence data available on the NCBI SRA database. Thus, there were 16 samples that were excluded from our analysis because four of them were not available on the SRA, eight did not have corresponding PCR and WGS *Wolbachia* infection statuses, and four had infection statuses determined by old data. See Supplementary Method S5 and Table S8 for subsampling validation methods and pipeline accuracy to divergent reference strains. See Supplementary Method S5 and Table S7 for a comparison of our approach to other methods [13].

Beta-binomial estimation of reproductive manipulator prevalence.

Beta-binomial distributions have been used to fit *Wolbachia* infectious status across species previously [46]. The beta-binomial model considers N random variables, X_i, which are all binomially distributed (i.e. infected vs. not-infected), but each with different parameters q_i and n_i, so that $X_i \sim Bin(q_i, n_i)$ (Figure S7). Using the approach developed in [46], we (1) determined moment estimators *u* and *s*, (2) beta distribution parameters α and β, and the global infection rate *x.* After we fit a beta distribution to the data, we took the integral from *c* to 1, where *c* is the minimum infection rate of a species to be considered positively infected. For example, a value of 0.001

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

means if one individual in 1000 is classified as positive, the species would be classified as being positively infected. For consistency with previous work, we used the value of 0.001 (as in [46]). See Supplementary Method S3, Figure S8, and Table S13 for beta-binomial rationale and downsampling results.

Estimating Endosymbiont Titer

We estimated the ratio of endosymbiont genome compliments to host genome compliments, hereafter referred to as titer. First, we computed the number of endosymbiont genome compliments through our BLAST-based method (Supplementary Method S1). Next, we computed the number of host genome compliments by aligning DNA sequencing reads to single copy orthologous arthropod proteins and taking the average of maximum depth across proteins with hits. We report the symbiont haploid : host pseudo haploid titer computation throughout this manuscript. See Supplementary Methods S7 and Figures S9 and S10 for validation of our approach to estimate host genome compliments.

Drosophila oocyte sampling, imaging, and analysis

We obtained *Drosophila melanogaster* and *Drosophila simulans* fly stocks infected with the wMel and wRi strains of *Wolbachia* from the Sullivan Lab. Oocytes from stage 9/10a flies were stained and mounted on glass slides and imaged with a SP5 Leica confocal microscope. We analyzed the fluorescence due to *Wolbachia* as described in [78] (See Supplementary Method S4).

Software Availability

The scripts used to classify microbial reproductive manipulator infections and estimate titer can be accessed through GitHub (www.github.com/pamedina/prevalence).

Acknowledgements

We thank the Corbett-Detig Lab and Sullivan Lab members for helpful feedback on this work. During this research, PM was supported by NIH training grant (T32 HG008345). This work was supported in part by an Alfred P. Sloan Fellowship and by NIH/NIGMS R35GM128932 to RC-D.

Contributions

Conceived and designed research: PM, RC-D, SLR. Performed the computational screen: PM. Analyzed sequence data and developed software: PM. Oocyte imaging and analysis: SLR. Wrote the manuscript: PM, SLR, RC-D. Edited the manuscript: RC-D, SLR, PM.

References

1. Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Bouletreau M. Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp. *Proc Natl Acad Sci U S A* 2001; **98**: 6247–6252.
2. Stevens L, Giordano R, Fialho RF. Male-Killing, Nematode Infections, Bacteriophage Infection, and Virulence of Cytoplasmic Bacteria in the GenusWolbachia. *Annu Rev Ecol Syst* 2001; **32**: 519–545.
3. Stouthamer R, Breeuwer JA, Hurst GD. Wolbachia pipientis: microbial manipulator of arthropod reproduction. *Annu Rev Microbiol* 1999; **53**: 71–102.
4. Werren JH. Biology of Wolbachia. *Annu Rev Entomol* 1997; **42**: 587–609.
5. Russell SL. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. *FEMS Microbiol Lett* 2019; **366**.
6. Bandi C, Dunn AM, Hurst GD, Rigaud T. Inherited microorganisms, sex-specific virulence and reproductive parasitism. *Trends Parasitol* 2001; **17**: 88–94.
7. Charlat S, Hurst GDD, Merçot H. Evolutionary consequences of Wolbachia infections. *Trends Genet* 2003; **19**: 217–223.

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8. Hurst GD, Werren JH. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* 2001; **2**: 597–606.
9. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. Wolbachia as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* 2010; **107**: 769–774.
10. Osborne SE, Leong YS, O'Neill SL, Johnson KN. Variation in antiviral protection mediated by different Wolbachia strains in Drosophila simulans. *PLoS Pathog* 2009; **5**: e1000656.
11. Yadav S, Frazer J, Banga A, Pruitt K, Harsh S, Jaenike J, et al. Endosymbiont-based immunity in Drosophila melanogaster against parasitic nematode infection. *PLoS One* 2018; **13**: e0192183.
12. Mazzetto F, Gonella E, Alma A, Others. Wolbachia infection affects female fecundity in Drosophila suzukii. *Bull Insectology* 2015; **68**: 153–157.
13. Pascar J, Chandler CH. A bioinformatics approach to identifying Wolbachia infections in arthropods. *PeerJ* 2018; **6**: e5486.
14. Jeyaprakash A, Hoy MA. Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Mol Biol* 2000; **9**: 393–405.
15. Baldo L, Ayoub NA, Hayashi CY, Russell JA, Stahlhut JK, Werren JH. Insight into the routes of Wolbachia invasion: high levels of horizontal transfer in the spider genus Agelenopsis revealed by Wolbachia strain and mitochondrial DNA diversity. *Mol Ecol* 2008; **17**: 557–569.
16. Augustinos AA, Santos-Garcia D, Dionyssopoulou E, Moreira M, Papapanagiotou A, Scarvelakis M, et al. Detection and characterization of Wolbachia infections in natural populations of aphids: is the hidden diversity fully unraveled? *PLoS One* 2011; **6**: e28695.
17. Cordaux R, Pichon S, Hatira HBA, Doublet V, Grève P, Marcadé I, et al. Widespread Wolbachia infection in terrestrial isopods and other crustaceans. *Zookeys* 2012; 123–131.
18. Simoes PM, Mialdea G, Reiss D, Sagot M-F, Charlat S. Wolbachia detection: an assessment of standard PCR protocols. *Mol Ecol Resour* 2011; **11**: 567–572.
19. Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, et al. The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. *BMC Biol* 2008; **6**: 27.
20. Serbus LR, Ferreccio A, Zhukova M, McMorris CL, Kiseleva E, Sullivan W. A feedback loop between Wolbachia and the Drosophila gurken mRNP complex influences Wolbachia titer. *J Cell Sci* 2011; **124**: 4299–4308.
21. Chong RA, Moran NA. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. *Proc Natl Acad Sci U S A* 2016; **113**: 13114–13119.
22. Funkhouser-Jones LJ, van Opstal EJ, Sharma A, Bordenstein SR. The Maternal Effect Gene Wds Controls Wolbachia Titer in Nasonia. *Curr Biol* 2018; **28**: 1692–1702.e6.
23. Mergaert P. Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat Prod Rep* 2018; **35**: 336–356.
24. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, et al. A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. *Cell* 2009; **139**: 1268–1278.
25. Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, Muzzi F, et al. Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature* 2011; **476**: 454–457.
26. Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, et al. The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *Nature* 2011; **476**: 450–453.
27. Frentiu FD, Zakir T, Walker T, Popovici J, Pyke AT, van den Hurk A, et al. Limited dengue virus replication in field-collected Aedes aegypti mosquitoes infected with Wolbachia. *PLoS Negl Trop Dis* 2014; **8**: e2688.
28. Zabalou S, Riegler M, Theodorakopoulou M, Stauffer C, Savakis C, Bourtzis K. Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. *Proc Natl Acad Sci U S A* 2004; **101**: 15042–15045.
29. Nikolouli K, Colinet H, Renault D, Enriquez T, Mouton L, Gibert P, et al. Sterile insect technique and Wolbachia symbiosis as potential tools for the control of the invasive species Drosophila suzukii. *J Pest Sci* 2018; **91**: 489–503.
30. Landmann F, Orsi GA, Loppin B, Sullivan W. Wolbachia-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathog* 2009; **5**: e1000343.
31. Bourtzis K, Nirgianaki A, Markakis G, Savakis C. Wolbachia infection and cytoplasmic incompatibility in Drosophila species. *Genetics* 1996; **144**: 1063–1073.

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

32. Leinonen R, Sugawara H, Shumway M, Collaboration INSD. The sequence read archive. *Nucleic Acids Res* 2010; **39**: D19–D21.
33. Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, et al. Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. *PLoS Genet* 2012; **8**: e1003129.
34. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. *Nature* 2012; **482**: 173–178.
35. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res* 2014; **24**: 1193–1208.
36. Gerth M, Gansauge M-T, Weigert A, Bleidorn C. Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic. *Nat Commun* 2014; **5**: 5117.
37. Wasala SK, Brown AMV, Kang J, Howe DK, Peetz AB, Zasada IA, et al. Variable Abundance and Distribution of Wolbachia and Cardinium Endosymbionts in Plant-Parasitic Nematode Field Populations. *Front Microbiol* 2019; **10**: 964.
38. Casiraghi M, Bain O, Guerrero R, Martin C, Pocacqua V, Gardner SL, et al. Mapping the presence of Wolbachia pipientis on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. *Int J Parasitol* 2004; **34**: 191–203.
39. Bandi C, Anderson TJ, Genchi C, Blaxter ML. Phylogeny of Wolbachia in filarial nematodes. *Proc Biol Sci* 1998; **265**: 2407–2413.
40. Casiraghi M, Anderson TJ, Bandi C, Bazzocchi C, Genchi C. A phylogenetic analysis of filarial nematodes: comparison with the phylogeny of Wolbachia endosymbionts. *Parasitology* 2001; **122 Pt 1**: 93–103.
41. Brown AMV, Wasala SK, Howe DK, Peetz AB, Zasada IA, Denver DR. Genomic evidence for plant-parasitic nematodes as the earliest Wolbachia hosts. *Sci Rep* 2016; **6**: 34955.
42. Zchori-Fein E, Perlman SJ. Distribution of the bacterial symbiont Cardinium in arthropods. *Mol Ecol* 2004; **13**: 2009–2016.
43. Russell JA, Funaro CF, Giraldo YM, Goldman-Huertas B, Suh D, Kronauer DJC, et al. A veritable menagerie of heritable bacteria from ants, butterflies, and beyond: broad molecular surveys and a systematic review. *PLoS One* 2012; **7**: e51027.
44. Sazama EJ, Ouellette SP, Wesner JS. Bacterial Endosymbionts Are Common Among, but not Necessarily Within, Insect Species. *Environ Entomol* 2019.
45. Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A, Weiss-Gayet M, et al. Antimicrobial peptides keep insect endosymbionts under control. *Science* 2011; **334**: 362–365.
46. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. How many species are infected with Wolbachia?--A statistical analysis of current data. *FEMS Microbiol Lett* 2008; **281**: 215–220.
47. Werren JH, Windsor DM. Wolbachia infection frequencies in insects: evidence of a global equilibrium? *Proc Biol Sci* 2000; **267**: 1277–1285.
48. Werren JH, Windsor D, Guo LR. Distribution of Wolbachia among neotropical arthropods. *Proc R Soc Lond B Biol Sci* 1995; **262**: 197–204.
49. West SA, Cook JM, Werren JH, Godfray HCJ. Wolbachia in two insect host--parasitoid communities. *Mol Ecol* 1998; **7**: 1457–1465.
50. Kikuchi Y, Fukatsu T. Diversity of Wolbachia endosymbionts in heteropteran bugs. *Appl Environ Microbiol* 2003; **69**: 6082–6090.
51. Nirgianaki A, Banks GK, Frohlich DR, Veneti Z, Braig HR, Miller TA, et al. Wolbachia infections of the whitefly Bemisia tabaci. *Curr Microbiol* 2003; **47**: 93–101.
52. Tagami Y, Miura K. Distribution and prevalence of Wolbachia in Japanese populations of Lepidoptera. *Insect Mol Biol* 2004; **13**: 359–364.
53. Gotoh T, Noda H, Hong X-Y. Wolbachia distribution and cytoplasmic incompatibility based on a survey of 42 spider mite species (Acari: Tetranychidae) in Japan. *Heredity* 2003; **91**: 208–216.
54. Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, et al. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc Natl Acad Sci U S A* 2015; **112**: 12764–12769.
55. Michonneau F, Brown JW, Winter DJ. rotl : an R package to interact with the Open Tree of Life data. *Methods Ecol Evol* 2016; **7**: 1476–1481.
56. Moreau J, Bertin A, Caubet Y, Rigaud T. Sexual selection in an isopod with Wolbachia-induced sex reversal: males prefer real females. *J Evol Biol* 2001; **14**: 388–394.

bioRxiv preprint doi: https://doi.org/10.1101/679837; this version posted June 23, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

57. McGraw EA, Merritt DJ, Droller JN, O'Neill SL. Wolbachia density and virulence attenuation after transfer into a novel host. *Proc Natl Acad Sci U S A* 2002; **99**: 2918–2923.
58. Duron O, Labbé P, Berticat C, Rousset F, Guillot S, Raymond M, et al. High Wolbachia density correlates with cost of infection for insecticide resistant Culex pipiens mosquitoes. *Evolution* 2006; **60**: 303–314.
59. Boyle L, O'Neill SL, Robertson HM, Karr TL. Interspecific and intraspecific horizontal transfer of Wolbachia in Drosophila. *Science* 1993; **260**: 1796–1799.
60. Reynolds KT, Hoffmann AA. Male age, host effects and the weak expression or non-expression of cytoplasmic incompatibility in Drosophila strains infected by maternally transmitted Wolbachia. *Genet Res* 2002; **80**: 79–87.
61. Veneti Z, Clark ME, Karr TL, Savakis C, Bourtzis K. Heads or tails: host-parasite interactions in the Drosophila-Wolbachia system. *Appl Environ Microbiol* 2004; **70**: 5366–5372.
62. Perrot-Minnot MJ, Werren, J. H. (Department of Biology, University of Rochester, Rochester, NY 14620 (USA)). Wolbachia infection and incompatibility dynamics in experimental selection lines. *J Evol Biol* 1999; **12**.
63. Enomoto S, Chari A, Clayton AL, Dale C. Quorum Sensing Attenuates Virulence in Sodalis praecaptivus. *Cell Host Microbe* 2017; **21**: 629–636.e5.
64. Yoder JB, Tiffin P. Sanctions, Partner Recognition, and Variation in Mutualism. *The American Naturalist* . 2017. , **190**: 491–505
65. Voronin D, Cook DAN, Steven A, Taylor MJ. Autophagy regulates Wolbachia populations across diverse symbiotic associations. *Proc Natl Acad Sci U S A* 2012; **109**: E1638–46.
66. Hurst GDD, Frost CL. Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harb Perspect Biol* 2015; **7**.
67. Correa CC, Ballard JWO. Wolbachia Associations with Insects: Winning or Losing Against a Master Manipulator. *Front Ecol Evol* 2016; **3**: 506.
68. Harumoto T, Anbutsu H, Lemaitre B, Fukatsu T. Male-killing symbiont damages host's dosage-compensated sex chromosome to induce embryonic apoptosis. *Nat Commun* 2016; **7**: 12781.
69. Hayashi M, Watanabe M, Yukuhiro F, Nomura M, Kageyama D. A Nightmare for Males? A Maternally Transmitted Male-Killing Bacterium and Strong Female Bias in a Green Lacewing Population. *PLoS One* 2016; **11**: e0155794.
70. Parratt SR, Frost CL, Schenkel MA, Rice A, Hurst GDD, King KC. Superparasitism Drives Heritable Symbiont Epidemiology and Host Sex Ratio in a Wasp. *PLoS Pathog* 2016; **12**: e1005629.
71. Pang R, Chen M, Yue L, Xing K, Li T, Kang K, et al. A distinct strain of Arsenophonus symbiont decreases insecticide resistance in its insect host. *PLoS Genet* 2018; **14**: e1007725.
72. Ryoo HD, Li J, Kang M-J. Drosophila XBP1 expression reporter marks cells under endoplasmic reticulum stress and with high protein secretory load. *PLoS One* 2013; **8**: e75774.
73. Fattouh N, Cazevieille C, Landmann F. Wolbachia endosymbionts subvert the endoplasmic reticulum to acquire host membranes without triggering ER stress. *PLoS Negl Trop Dis* 2019; **13**: e0007218.
74. Landmann F, Bain O, Martin C, Uni S, Taylor MJ, Sullivan W. Both asymmetric mitotic segregation and cell-to-cell invasion are required for stable germline transmission of Wolbachia in filarial nematodes. *Biology Open* . 2012. , **1**: 536–547
75. Sheehan KB, Martin M, Lesser CF, Isberg RR, Newton ILG. Identification and Characterization of a Candidate Wolbachia pipientis Type IV Effector That Interacts with the Actin Cytoskeleton. *MBio* 2016; **7**.
76. Newton ILG, Savytskyy O, Sheehan KB. Wolbachia utilize host actin for efficient maternal transmission in Drosophila melanogaster. *PLoS Pathog* 2015; **11**: e1004798.
77. White PM, Serbus LR, Debec A, Codina A, Bray W, Guichet A, et al. Reliance of Wolbachia on High Rates of Host Proteolysis Revealed by a Genome-Wide RNAi Screen of Drosophila Cells. *Genetics* 2017; **205**: 1473–1488.
78. Russell SL, Lemseffer N, White PM, Sullivan WT. Wolbachia and host germline components compete for kinesin-mediated transport to the posterior pole of the Drosophila oocyte. *PLoS Pathog* 2018; **14**: e1007216.