Extensive genomic diversity of closely related Wolbachia strains

Nadeeza Ishmael,1,2† Julie C. Dunning Hotopp,1,2† Panagiotis Ioannidis,3 Sarah Biber,4§ Joyce Sakamoto,2 Stefanos Siozios,3 Vishvanath Nene,1,2|| John Werren,5 Kostas Bourtzis,3 Seth R. Bordenstein4,6 and Hervé Tettelin1,2

1 J. Craig Venter Institute, 9708 Medical Center Dr., Rockville, MD 20850, USA
2 Institute for Genome Sciences, Department of Microbiology and Immunology, University of Maryland Baltimore, Baltimore, MD 21201, USA
3 Department of Environmental and Natural Resources Management, University of Ioannina, 30100 Agrinio, Greece
4 Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, The Marine Biological Laboratory, Woods Hole, MA 02543, USA
5 Biology Department, University of Rochester, Rochester, NY 14627, USA
6 Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

Using microarray-based comparative genome hybridization (mCGH), the genomic content of Wolbachia pipientis wMel from Drosophila melanogaster was compared to the closely related Wolbachia from D. innubila (wInn), D. santomea (wSan), and three strains from D. simulans (wAu, wRi, wSim). A large number of auxiliary genes are identified in these five strains, with most absent/divergent genes being unique to a given strain. Each strain caused an average of ~60 genes to be removed from the core genome. As such, these organisms do not appear to have the streamlined genomes expected of obligate intracellular bacteria. Prophage, hypothetical and ankyrin repeat genes are over-represented in the absent/divergent genes, with 21–87% of absent/divergent genes coming from prophage regions. The only wMel region absent/divergent in all five query strains is that containing WD_0509 to WD_0511, including a DNA mismatch repair protein MutL-2, a degenerate RNase, and a conserved hypothetical protein. A region flanked by the two portions of the WO-B prophage in wMel is found in four of the five Wolbachia strains as well as on a plasmid of a rickettsial endosymbiont of Ixodes scapularis, suggesting lateral gene transfer between these two obligate intracellular species. Overall, these insect-associated Wolbachia have highly mosaic genomes, with lateral gene transfer playing an important role in their diversity and evolution.

INTRODUCTION

Wolbachia bacteria are common obligate intracellular endosymbionts that infect a wide variety of invertebrates, including arthropods and filarial nematodes. The arthropod-infecting Wolbachia strains are maternally inherited and exert unusual effects on host reproduction, including: (1) parthenogenesis, whereby infected virgin females produce infected female offspring, (2) male killing, whereby infected male embryos fail to develop, (3) feminization, whereby genetic males develop into reproductively capable females, and (4) cytoplasmic incompatibility, the most common phenotype, whereby the offspring of uninfected females and infected males fail to develop (Stouthamer et al., 1999; Werren et al., 2008). These phenotypes increase the number of infected hosts within a population, promoting the frequency of Wolbachia. Although insect-associated Wolbachia are typically considered reproductive parasites, some
Wolbachia strains have no reproductive phenotype (Hoffmann et al., 1996; Turelli & Hoffmann, 1995), and mutualistic host–bacteria interactions have also been described (e.g. Dobson et al., 2002; Hedges et al., 2008; Teixera et al., 2008). Phylogenetic analysis of concatenated protein sequences and the 16S rRNA places Wolbachia in a clade of obligate intracellular bacteria including the tick- and trematode-vectored vertebrate pathogens from the genera Anaplasm, Ehrlichia and Neorickettsia, as well as the more distantly related, but better characterized, arthropod-associated Rickettsia spp. (Dumler et al., 2001; Dunning Hotopp et al., 2006b). Numerous phylogenetic analyses of Wolbachia have been undertaken, including single-locus and multi-locus approaches (Bordenstein et al., 2008; Casiraghi et al., 2005; Lo et al., 2002; O’Neill et al., 1992; Werren et al., 1995; Zhou et al., 1998). Despite maternal inheritance in arthropods, the phylogenetic distribution of these Wolbachia is not concordant with that of their hosts (Baldo et al., 2006b; O’Neill et al., 1992; Werren et al., 1995). Therefore, in arthropods both horizontal transmission of Wolbachia and loss of infection are likely, although the mechanisms are not fully understood (Bouchon et al., 1998; Heath et al., 1999; Huigens et al., 2004).

Species designations in Wolbachia have remained elusive, largely owing to the lack of an accepted definition of bacterial species that can be applied to obligate intracellular bacteria. However, current phylogenetic and recombination analysis supports all Wolbachia strains being formally included in the single species Wolbachia pipientis (Lo et al., 2007). To help distinguish isolates, Wolbachia strains have been divided into eight phylogenetic clades termed supergroups (Bordenstein et al., 2008; Lo et al., 2007). The A and B supergroup Wolbachia are the most common and are widespread in arthropods as single or double infections (Werren et al., 1995). The C and D supergroups contain the Wolbachia strains from filarial nematodes (Lo et al., 2002). Primitive wingless insects (springtails; Collembola) are infected with supergroup E (Czarnecki & Tebbe, 2004; Lo et al., 2002; Vandekerckhove et al., 1999). Supergroup F consists of Wolbachia infecting termites, weevils, true bugs and filarial nematodes (Casiraghi et al., 2001; Lo et al., 2002; Rasgon & Scott, 2004). Supergroup H infects termites (Bordenstein & Rosengaus, 2005). Three other divergent lineages exist that have not been given supergroup designations, including the infection found in the flea Ctenocephalides canis (Casiraghi et al., 2004), the filarial nematode Dipetalonema gracile (Casiraghi et al., 2004), and the pseudoscorpion Cordylochernes scorpioides (Zeh et al., 2005; Zeh & Zeh, 2006). Double or triple infections have been described with isolates coming from different or identical supergroups (Werren et al., 1995).

Obligate intracellular bacteria are often viewed as excellent model systems to understand the stable, symbiotic interactions that occur between prokaryotes and their eukaryotic hosts. However, recent analyses of sequenced genomes of host-switching obligate intracellular bacteria, like Wolbachia, suggest that mobile elements and genetic exchange are not as rare as previously thought (Bordenstein & Reznikoff, 2005). The presence of extensive recombination (Baldo et al., 2006a; Klasson et al., 2009), mobile elements (Klasson et al., 2008; Wu et al., 2004) and lateral phage transfer between co-infecting strains (Bordenstein & Werngreen, 2004; Masui et al., 2000) facilitates high levels of genetic exchange in Wolbachia. These processes that shape the Wolbachia genome are likely influenced by exposure to novel gene pools available due to horizontal transmission of infection, multiple infections and large numbers of infected hosts.

The high level of genetic exchange in Wolbachia suggests that like many free-living bacteria, Wolbachia has a core genome complemented by an extensive auxiliary genome. A core genome consists of all the genes shared by all (or almost all) sequenced strains for a given taxon, typically a species (Boucher et al., 2001; Hogg et al., 2007; Lan & Reeves, 2000; Tettelin et al., 2005, 2008). All of the other genes constitute what has been termed the dispensable (Tettelin et al., 2005, 2008), contingency (Hogg et al., 2007) or auxiliary (Boucher et al., 2001; Lan & Reeves, 2000) genome, which encompasses the genetic variation within the species. A subset of the auxiliary genes includes the unique (Hogg et al., 2007) or strain-specific genes (Tettelin et al., 2005, 2008), those genes found only in a single genome of the species. When the auxiliary genome is combined with the core genome, the result is the pan-genome (Tettelin et al., 2005, 2008) or supragenome (Hogg et al., 2007). Pan-genomes can be classed as open/unbounded or closed/bounded (Medini et al., 2005; Tettelin et al., 2008). With an open pan-genome, each additional genome sequenced, on average, contributes new genes to the pan-genome. In contrast, with a closed pan-genome, each additional genome sequenced contributes no new genes to the pan-genome. Species with open pan-genomes include Streptococcus agalactiae, Streptococcus pneumoniae, Escherichia coli, Bacillus cereus and Pseudomonas marinus (Tettelin et al., 2008). Those with closed pan-genomes include Staphylococcus aureus, Streptococcus pyogenes, Ureaplasma urealyticum and Bacillus anthracis (Tettelin et al., 2008). In addition to the above nomenclature, another convention has been proposed that combines all genes found in >1% of strains into a species genome, and any genes in <1% of strains are considered of foreign origin or on the decline (Boucher et al., 2001; Lan & Reeves, 2000).

In order to examine the core genome and the genetic flux between W. pipientis strains, we examined the genomic content of five strains using microarray-based comparative genomic hybridization (mCGH) on a Wolbachia microarray. DNAs from reference and query organisms are differentially labelled and competitively hybridized to a microarray. The hybridization data reveal the presence/absence of genes in the query genome relative to the genome content of the reference organism. For Wolbachia, a large number of auxiliary genes are
identified in five closely related supergroup A strains, with most absent/divergent genes unique to a given strain. Prophage, hypothetical and ankyrin repeat genes are over-represented in the absent/divergent genes, with 3–55 of those being unique to a given strain. One variable region flanked by the two portions of the WO-B prophage in wMel is also identified in a plasmid of a Rickettsia endosymbiont of Ixodes scapularis (Noda et al., 1997) suggesting that lateral gene transfer extends across the order Rickettsiales.

**METHODS**

**Bacterial strains.** Genomic DNA from six Wolbachia strains was hybridized in this study (Table 1). These are a subset of well-studied supergroup A strains that were chosen because they were expected to be genetically similar but induce a variety of reproductive phenotypes. The supergroup A Wolbachia strains from Nasonia vitripennis 12.1, Drosophila recondens, D. yakuba and D. teissieri were also initially chosen for inclusion, but were later removed because of difficulties in background hybridization, likely due to host DNA contamination.

**Microarray construction.** A wMel microarray was constructed which represents 95% of the unique ORFs (Wu et al., 2004). Primers were designed from the unique predicted genes annotated in wMel. In order to obtain pure amplicons devoid of background amplification of host genes, all amplicons were produced, where possible, from two separate clones from the wMel genome-sequencing project. Genes in physical gaps where a sequencing clone was not available were amplified directly from genomic DNA from Drosophila melanogaster infected with wMel. Amplicons (50–1000 bp) were produced using AmpliTaq (Applied Biosystems), purified using filtration plates (Millipore), and analysed on agarose gels. The purified amplicons were diluted 1:1 with DMSO, spotted in quadruplicate onto Corning UltraGap slides with a Lucidea printing robot (GE Healthcare) and irradiated with UV light (Dunning Hotopp et al., 2006a).

**Genomic DNA preparation and whole-genome amplification.** Wolbachia cells were isolated from 2–4 g of live, whole insects via density centrifugation, and whole chromosomal fragments were purified following pulsed-field gel separation as described by Wernegreen et al. (2002). Due to the small amounts of chromosomal fragments recovered, whole-genome amplification was needed to produce enough DNA for hybridization. Therefore ~10 ng of genomic DNA was amplified per reaction with GenomiPhi. A similar strategy was used successfully for mcGH studies in Anaplasma phagocytophilum (Dunning Hotopp et al., 2006b). Whole-genome amplification has been shown to be specific, with a randomly distributed bias and a mutation rate of <2 × 10⁻³ (Zhang et al., 2006). In the obligate intracellular bacterium Ehrlichia chaffeensis Arkansas, a comparison of genomic DNA with GenomiPhi-amplified DNA also revealed a randomly distributed bias (data not shown). A similar comparison of genomic DNA to GenomiPhi-amplified DNA was not possible for Wolbachia DNA in the current study due to an inability to obtain large quantities of genomic DNA. However, a comparison of two independent GenomiPhi amplifications of the same genomic DNA sample from strain wRi demonstrates the reproducibility of the amplification (Fig. 1). To reduce the effects of any bias, each probe was synthesized from a pool of three distinct GenomiPhi amplifications.

**Microarray hybridizations.** Cy3 and Cy5 probes were synthesized from GenomiPhi-amplified DNAs as described by Tettelin et al. (2001). Briefly, amino-allyl-dUTP-labelled probes were synthesized from 4 µg of genomic DNA using Klenow fragment (3’→5’ exo-) (New England Biolabs). The reactions were purified using the QiAquick PCR purification kit (Qiagen) with modified phosphate buffers. Cy3 or Cy5 dyes (GE Healthcare) were chemically coupled to the incorporated amino-allyl-dUTP in carbonate buffer. Probe pairs were resuspended in 5× SSC, 50 % formamide, 0.1 % SDS, 0.1 mM DTT and 0.35 mg ml⁻¹ salmon sperm DNA and hybridized to slides overnight at 42 °C. The hybridized slides were washed twice in 2× SSC, 0.1 % SDS at 55 °C; twice in 0.1× SSC, 0.1 % SDS at room temperature; twice in 0.1× SSC at room temperature; and once in MilliQ water at room temperature. They were then dried, and scanned using a GenePix4000B scanner (Molecular Devices).

**Microarray data analysis and bioinformatics.** Cy3- and Cy5-labelled probes were synthesized from a pool of three distinct GenomiPhi amplifications to reduce bias. One flip-dye experiment (two hybridizations) was performed with each of three independent pools, yielding a total of six hybridizations per strain. With 4–8 printed replicates per slide this yielded 24–48 replicated spots per gene to compare across the study. Ratios were calculated as wMel reference fluorescence over the query fluorescence (reference/query) and were normalized using iterative log-mode centring (Dunning Hotopp et al., 2006a). The geometric mean ratio was calculated for all good spots in each flip-dye experiment. The values reported are the medians of the geometric mean ratio for the three flip-dye experiments. Ideally, hybridization results would be compared to gene presence/absence from two closed genome sequences to establish microarray platform-specific criteria for absence/divergence. However, only one supergroup A Wolbachia has had its genome fully sequenced and closed. Therefore, previously established criteria for a Neisseria meningitidis microarray were used (Dunning Hotopp et al., 2006a), with genes deemed absent/divergent (ratio >3), present (ratio <2), or not predictive (ratio 2–3).

**Southern hybridizations.** Total genomic DNA from Drosophila lines was extracted according to the Holmes–Boner method (Sambrook et al., 1989). The DNA fragments generated from 5 µg

<table>
<thead>
<tr>
<th>Table 1. Wolbachia strains used in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strain</strong></td>
</tr>
<tr>
<td>wMel</td>
</tr>
<tr>
<td>wAu</td>
</tr>
<tr>
<td>wRi</td>
</tr>
<tr>
<td>wSim</td>
</tr>
<tr>
<td>wSan</td>
</tr>
</tbody>
</table>
Mobile genetic elements

Bacteriophage and other mobile elements are notable catalysts of lateral gene transfer and traditionally have been considered absent in the genomes of obligate intracellular bacteria. However, in species that experience horizontal transmission to new hosts, moderate to high levels of mobile genetic elements have been observed (Bordenstein & Reznikoff, 2005; Wu et al., 2004). Previous studies of Wolbachia phage indicate rampant lateral phage transfer between divergent strains and that the mechanism of transfer is in part mediated via the high rate of Wolbachia co-infections (Bordenstein & Wernegreen, 2004; Masui et al., 2000). Mobility of Wolbachia phage could promote the introduction of novel genes by shuttling large fragments of Wolbachia DNA between genomes.

The mCGH results show that phage genes are significantly over-represented in the absent/divergent genes (Table 2), supporting the idea that these strains with similar genomic backbones are actively exchanging phage regions. Of the absent/divergent genes identified, 21–87% are in the prophage regions (Fig. 3). Although all of the strains examined here have portions of each prophage (Fig. 4B), each also appears to be missing sections of these regions.

One region of interest is the region encompassing WD_0612–WD_0618 that is flanked by portions of wMel DNA between genomes. It encodes an NAD-dependent epimerase/dehydratase family protein, a glycosyltransferase with a radical SAM domain, two putative l-allo-threonine aldolases, an ABC transporter permease and ATP-binding protein, a conserved domain protein, and a conserved hypothetical protein (Fig. 4A). This region was conserved in all but wMel in the current study (Fig. 4B), but was not identified in this study, each query strain leads to the removal of ~60 genes on average from the core W. pipiensis genome (Fig. 2B). An exponential regression suggests that the core genome consists of 621 genes. Although this estimate is based on only five data points, it is roughly consistent with the 743 wMel genes in the 637 shared clusters of orthologous gene clusters between wMel and the most distantly related sequenced supergroup D Wolbachia genome, wBm, as described in the Rickettsiales Comparative Database (Dunning Hotopp et al., 2006b). The similarity of these numbers is somewhat surprising considering the similarity of the five strains compared by mCGH when compared to the divergence of wBm. In addition, the microarray-defined core genome uses high-stringency DNA identity as opposed to the bi-directional reciprocal best protein matches for the orthologue prediction. Despite these differences, 597 wMel proteins are present in both the orthologue prediction and the microarray-defined core genome. The convergence of these two core genome metrics across diverse strains may suggest a single Wolbachia core genome that spans the supergroup designations, thus supporting the single species designation of W. pipiensis (Lo et al., 2007).
in either the wBm or wPip genome sequences. Interestingly, this same region was uncovered in a bacterial plasmid in the genome sequencing of I. scapularis (GenBank ABJB000000000) and is attributed to a concomitantly sequenced ovary-associated Rickettsia endosymbiont (Noda et al., 1997; Steiner et al., 2008). This endosymbiont is most closely related to other tick-associated Rickettsia spp. (Weinert et al., 2009). The plasmid is not thought to arise from a lower-level contaminating Wolbachia infection because (a) only 44 Wolbachia sequences limited to this ~10 kbp region were recovered in the whole-genome shotgun sequencing project (Dunning Hotopp et al., 2007) as opposed to the 100 000+ reads recovered with homology to the rickettsial endosymbiont; (b) the 16S rRNA and wsp genes could not be amplified from the I. scapularis genomic DNA with Wolbachia-specific primers; and (c) of the 40 non-transposase ORFs on the plasmid, 23 have a rickettsial best match (e-value \(10^{-15}\)), including genes involved in plasmid maintenance, while only the seven corresponding to WD_0612–WD_0618 have a Wolbachia best match (e-value \(10^{-6}\)). Of the remaining non-transposase ORFs on the plasmid, eight have a Lawsonia intracellularis best match (e-value \(10^{-5}\)), and two have a best match to other bacteria.

In the Rickettsia endosymbiont plasmid, this ~10 kbp region is flanked by transposases (Fig. 4A). The lack of divergence (~70 % nt identity) and association with mobile elements strongly suggests lateral gene transfer between these micro-organisms. No closely related orthologues are in public databases to facilitate assigning the direction of this transfer or determining whether a third, intermediate organism was involved. Regardless, this is a compelling case of lateral gene transfer between intracellular bacteria of different genera.

As an additional example of mobile element transfer between intracellular bacteria of diverse genera, Wolbachia and the unrelated bacterial agent of cat-scratch disease, Bartonella henselae, share divergent prophage homologues of prophage WO-B (Alsmark et al., 2004). These two intracellular parasites likely exchanged phage genes in the

### Table 2. TIGR role categories of overrepresented absent/divergent genes

<table>
<thead>
<tr>
<th>Strain</th>
<th>TIGR role category (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosophila simulans wAu</td>
<td>Hypothetical proteins (1.3 \times 10^{-17})</td>
</tr>
<tr>
<td>Drosophila innabila wlnn</td>
<td>Hypothetical proteins (8.0 \times 10^{-3})</td>
</tr>
<tr>
<td>Drosophila simulans wRi</td>
<td>Hypothetical proteins (5.2 \times 10^{-6})</td>
</tr>
<tr>
<td></td>
<td>Prophage functions (5.6 \times 10^{-4})</td>
</tr>
<tr>
<td>Drosophila simulans wSim</td>
<td>Prophage functions (9.3 \times 10^{-3})</td>
</tr>
<tr>
<td>Drosophila simulans wSan</td>
<td>Prophage functions (2.7 \times 10^{-10})</td>
</tr>
<tr>
<td>Combined (all five strains)</td>
<td>Hypothetical proteins (3.9 \times 10^{-26})</td>
</tr>
<tr>
<td></td>
<td>Prophage functions (1.0 \times 10^{-4})</td>
</tr>
</tbody>
</table>

---

Fig. 2. Genomic diversity and the core genome. (A) Most gene differences are unique to a strain, resulting in different sets of absent/divergent genes (ratios \(\geq 3\)) per strain, with only three genes absent/divergent in all five strains. Numbers within the intersections of different ovals count the number of absent/divergent genes (ratio \(\geq 3\)) shared by 2, 3, 4 or 5 organisms. (B) The number of Wolbachia core genes for n query genomes is plotted against the number of query genomes compared. For example, when n=3, each query genome is compared to all possible combinations of two other query genomes to see how many genes are shared between the three. Each combination (●), the mean of each combination (■), and the standard deviation (error bars) are presented. With the addition of each genome a significant number of genes are removed from the core genome, demonstrating that even between these closely related Wolbachia strains there is a sizeable amount of genomic diversity. The nonlinear regression of \(620.9+557.9e^{-0.204x}\) (RSS=0.0148) suggests that the core genome consists of 621 genes.
Fig. 3. MLST-based phylogeny of mCGH strains. The phylogenetic relationship is shown of the mCGH strains (bold) based on MLST (figure adapted from Baldo et al., 2006b). The ST-13 complex is boxed in grey. This tree is contrasted to the total number of absent/divergent genes and the absent/divergent genes in the vicinity of prophages WO-A and WO-B. Since exact prophage boundaries cannot be discerned, for this analysis WO-A and WO-B are defined as the regions from WD_0250 to WD_0300 and WD_0581 to WD_0650, respectively. Large portions of the absent/divergent genes are found in the vicinity of the prophage. Despite close relationships among strains based on MLST, mCGH reveals significant levels of genome divergence among them.

Fig. 4. Prophage regions. Four prophage regions are compared, including the canonical wKue WO phage, the two wMel WO prophage, and a representative prophage from wPip (A). Absent/divergent genes (red; ratio >3), ambiguous genes (blue; ratio 2–3) and possible duplicated genes (green; ratio <0.3) are shown under the representative wMel prophage region (B). The prophage regions appear to have a high level of variation. A region flanked by wMel's WO-B is homologous to a plasmid sequenced from a rickettsial endosymbiont of I. scapularis (A). This region appears to be the result of lateral gene transfer between Wolbachia and Rickettsia. It is present in wAu, wSim, wRi and wSan in the current study, but not in the genomes of wBm or wPip (B).
evolutionary past within their common invertebrate host, the cat flea Ctenophalides felis (Casiraghi et al., 2005; Rolain et al., 2003).

Ankyrin repeat proteins and confirmation of mCGH results

The sequencing and annotation of the wMel genome revealed 23 genes encoding ankyrin domain repeats, one of the highest levels in a prokaryotic genome (Wu et al., 2004). The functions of these proteins in wMel are not known, although ankyrin repeats often promote protein–protein interactions. In eukaryotes, they can be involved in many functions, including cell cycle regulation and cytoskeleton interactions (Voronin & Kiseleva, 2007; Wu et al., 2004), making them candidates for mediating Wolbachia–host interaction and host reproductive phenotypes (Foster et al., 2005; Iturbe-Ormaetxe et al., 2005a). In other symbiotic relationships, ankyrin repeat genes have been shown to bind to host chromatin in A. phagocytophilum (Caturegli et al., 2000; Park et al., 2004), have a role in hydrogen peroxide resistance of Pseudomonas aeruginosa (Howell et al., 2000) and be key in establishing a plant–bacterium relationship between Lotus japonicus and Mesorhizobium loti (Kumagai et al., 2007).

The ankyrin protein content of nine Wolbachia strains was determined previously using hybridization and PCR amplification (Iturbe-Ormaetxe et al., 2005a). A comparison of the two strains overlapping with this current study reveals a high degree of agreement. Using a ratio cutoff of 3 and with the exception of WD_0633, the results for strain wAu are in agreement with those previous results, revealing that all of the examined ankyrin genes from wMel are present in wAu. In strain wRi, three genes (WD_0566, WD_0147, WD_0754) were previously demonstrated as being present in wRi but are suggested to be absent/divergent in the current mCGH study. There are several possible explanations for these differences, including: (1) size and sequence of the hybridization probe/target pairs in the two studies, (2) sensitivity, particularly pertaining to single- versus dual-channel hybridization experiments, (3) cutoff criteria for assessing background, (4) data normalization, (5) hybridization stringency, or (6) polymorphic populations. However, despite these differences there is a high degree of concordance in the dataset.

In the current study, only wRi and wSim had an over-representation of absent/divergent ankyrin genes (Fisher’s exact text, P<0.01). Phage activity, along with the movement of transposable genetic elements, may help to explain the marked changes in the ankyrin gene content, as most absent/divergent ankyrin repeat genes were localized to the prophage WO-A in these two genomes. However, this does not reflect an absence of the prophage, as WO-A structural genes and some conserved hypothetical genes are conserved (Fig. 4B). In contrast, wSan contains the ankyrin repeat proteins in WO-A, but has missing/divergent phage components (Fig. 4B). Instead of prophage absence, this likely reflects either differential loss of portions of the prophage region after incorporation into the genome and/or independent acquisition of the ankyrin genes. There appears to be a similar, but lower-level, variation in the WO-B ankyrin repeat genes in wSan and wRi (Fig. 4B), suggesting the presence of related prophage regions in Wolbachia strains with differing content of ankyrin repeat proteins.

Hypothetical proteins

Hypothetical proteins are also overrepresented in the absent/divergent genes in this study (Table 2). There are 72 hypothetical proteins ranging in size from 30 to 2843 aa with a ratio >=3, suggesting absence/divergence in at least one strain. All of these proteins were re-examined in order to improve annotation. There are 19 (26%) that are conserved with wPip, but 66 (92%) remain hypothetical proteins unique to arthropod Wolbachia. Of the remaining six genes, three have orthologues in the W. pipiens wBm genome (WD_0424, WD_0434, WD_1039), one has an orthologue in the Rickettsia felis genome (WD_0268), one has an orthologue in Ehrlichia chaffeensis (WD_0971), and one has orthologues in Anopheles gambiae and Aedes aegypti (WD_0513). A third of the absent/divergent hypothetical proteins are >150 aa, while the mean and median sizes are 209 aa and 68 aa, respectively. By comparison, the hypothetical proteins in the entire genome have the same size range but with mean and median sizes of 180 aa and 95 aa, respectively. Therefore, it is not merely the smallest hypothetical proteins, which one might hypothesize are incorrectly predicted to encode proteins, that are absent/divergent. As the unique genes in the genome, those encoding hypothetical proteins may reflect the genetic diversity of Wolbachia.

Divergence of strains

Most gene differences were unique to a strain, resulting in a unique set of absent/divergent genes per strain (Fig. 2A). Only three genes were present in wMel but absent/divergent in all other strains examined. These adjacent genes include a DNA mismatch repair protein MutL-2 (WD_0509), a degenerate RNase (WD_0510), and a conserved hypothetical protein (WD_0511). The region from WD_0509 to WD_0514 was previously hypothesized to be horizontally transferred into wMel based on absence in the Wolbachia draft genomes of wAna and wSim (Salzberg et al., 2005). In the current study, WD_0512 and WD_0513 are also absent/divergent in wAu and showed higher than average ratios in all strains but wSim. Using PCR and Southern hybridization, WD_0509–WD_0514 were found to be absent in wAu and wRi while only partially absent in wSan, confirming the mCGH studies. In wSim, these WD_0512 and WD_0513 genes actually showed lower than average ratios, suggesting multiple copies. However, these regions are absent in the D. simulans genome sequencing project and corresponding
wSim Wolbachia draft genome (Salzberg et al., 2005). These discrepancies may be attributed to a cloning bias in the whole-genome shotgun sequencing project or a Wolbachia population polymorphic for this region. Although unusual, a polymorphic population has some precedent. In the sequencing of wPip, the laboratory, inbred mosquito line used was not clonal, leading to the suggestion that Wolbachia can rapidly generate chromosomal variation (Klasson et al., 2008).

WD_0513 is interesting as its closest putative homologues are found in the genomes of Anopheles gambiae, Aedes aegypti and Culex quinquefasciatus (Woolfit et al., 2009) and facilitate invasion of malarial sporozoites into salivary gland cells (Arca et al., 2005; Korochkina et al., 2006). Prior analyses revealed that the co-transcribed WD_0512–WD_0514 genes (Iturbe-Ormaetxe et al., 2005a) are frequently absent from Wolbachia genomes with the exception of wPip (Korochkina et al., 2006) and Wolbachia closely related to wMel including wMelCS and wMelPop (Iturbe-Ormaetxe et al., 2005a). The limited distribution of these proteins in only mosquitoes and select Wolbachia strains suggests host/endosymbiont lateral gene transfer. Lateral transfer of Wolbachia genes to animal hosts has been described (Dunning Hotopp et al., 2007; Kondo et al., 2002), but in the case of WD_0513, the actual direction of this transfer is difficult to ascertain without further knowledge of the gene’s phylogenetic distribution (Woolfit et al., 2009). However, the limited distribution of this region and the presence of these proteins in both the endosymbiont and host make these genes interesting candidates for research into host–parasite interactions.

Although all of the strains examined here had distinct genomic content, wRi and wSim were found to be most similar (Figs 2A, 3 and 5). It has been hypothesized that

---

**Fig. 5.** Circular representation of mCGH results. The outer two rims present the predicted coding regions on the plus and minus strand, respectively, colour-coded by role categories. Rim 3 contains the genes not represented on the array (orange) and atypical nucleotide content (black). Rim 4 indicates the origin of replication (gold), phage genes (pink), and transposon genes (green). Rims 5–9 are the comparative genome hybridization results for wAu, wInn, wRi, wSim and wSan, respectively. Using previously determined cutoffs, the following assignments were made: a ratio >3 is most likely absent or divergent in the test strain (red); a ratio between 2 and 3 is absence/divergence not predicted (blue); a ratio <2 is most likely present in the test strain (white). Considerable diversity is present in the absent/divergent genes, particularly in the vicinity of phage (rim 4, pink).
these are actually the same strain (Iturbe-Ormaetxe et al., 2005b), and the two strains are identical in their MLST alleles (Fig. 3). In this current study, the divergence of strains could not be determined using microarray support trees analysis as all nodes had <50% support (TMeV support trees, average linkage, Euclidean distance, bootstrap 100 iterations; data not shown). However, these strains had the highest number of absent/divergent genes in common in all of the two-way comparisons (Fig. 2A). Despite this commonality, there are 46 differences in absent/divergent genes between the two strains while only 21 absent/divergent genes overlapped in these two strains (Fig. 2A), suggesting distinct genomic content.

**DISCUSSION**

**Genome diversity**

Although Wolbachia are obligate intracellular bacteria, they are considered facultative endosymbionts because hosts can survive uninfected. They are cosmopolitan, with host promiscuity and frequent co-infection with other Wolbachia strains and probably other intracellular bacteria. These factors combined likely increase the opportunity of Wolbachia to exchange genomic DNA. This increased genomic exchange may promote genetic and phenotypic diversity and contribute to Wolbachia's success as one of the most abundant intracellular bacteria.

**Lateral gene transfer**

In the supergroup A and B Wolbachia, intraspecies lateral gene transfer is facilitated by prevalent infections (Hilgenboecker et al., 2008), transfer of mobile elements (Bordenstein & Wernegreen, 2004; Masui et al., 2000) and extensive recombination (Baldo et al., 2006a, b; Lo et al., 2007; Verne et al., 2007). It has not been clear if Wolbachia can exchange DNA through lateral gene transfer with other bacteria. In order for this to happen, Wolbachia must come into contact with other bacteria either inside the cell or as it is passing between cells/organisms (e.g. in the hemolymph). We show here a region shared at ~70% nucleotide identity flanked by transposases on a plasmid in a Rickettsia endosymbiont and flanked by prophage sequences in a Wolbachia. The direction of the transfer is difficult to identify. This region is not found in all Rickettsia spp. or all Wolbachia strains, and has no close similarity to other organisms to allow a phylogenetic assignment. Both organisms can infect ovaries and can infect I. scapularis (Benson et al., 2004), possibly providing the opportunity for plasmid transfer. Alternatively, another host or a third DNA reservoir may be involved in mediating this transfer.

**Limitations/benefits of mCGH**

mCGH is of limited use in genomic comparison of intracellular bacteria with low genomic diversity. For example, a comparison of two A. phagocytophilum genomes revealed that only four p44 genes varied between the strains (Dunning Hotopp et al., 2006b). These encompass a subset of genes that were already known to be highly variable. But in the case of Wolbachia, where large genomic variation is seen, mCGH is more fruitful.

The mCGH platform may become obsolete as sequencing costs fall, but with in-house printed arrays the cost is still an order of magnitude less than most commercial microarray platforms and two orders of magnitude less than data obtained through next-generation sequencing. So mCGH may still have significant use in assessing genomic diversity, particularly before selecting strains to be sequenced.

One problem with examining obligate intracellular bacteria with any genomic platform is acquiring microgram quantities of DNA that are devoid of host genomic DNA contamination. Better methods are needed to obtain purified material and to amplify materials when genomic DNA quantity is limited. In this current study, several strains had to be eliminated due to high levels of background hybridization, likely owing to host DNA contamination. Likewise, because of limiting quantities of highly purified Wolbachia DNA, whole-genome amplification could not be avoided. These limiting DNA quantities precluded assessment of bias in the current study, but the amplification results were shown to be reproducible. By pooling multiple amplification reactions, performing multiple replicate hybridizations, and taking the median value of those replicates any biases should be minimized.

One of the limitations of mCGH, when compared to sequencing, is that it only assesses gene content represented on the microarray. As such, only sequences represented in public databases are queried. Although great genomic diversity of wMel prophage regions is demonstrated, novel phage regions and genes cannot be assessed using the mCGH technique. The high levels of variation in these regions suggest that further sequencing of Wolbachia and its phage would be particularly fruitful.

**Concluding remarks**

In Wolbachia, mCGH can identify variable regions including phage regions, ankyrin repeat proteins, and hypothetical proteins. However, one of the limitations is the ability to only characterize the diversity relative to known gene content. The unique genes in each strain examined remain elusive. But even with this limitation, the difference in gene content is the greatest variation observed for closely related obligate intracellular bacteria, with Wolbachia phages being a major source of variation. Wolbachia sequencing has been stymied due to the high levels of host DNA contamination. But as the costs of genome sequencing decrease and new sequencing technologies are developed that alleviate this burden, Wolbachia strains should be targeted for further sequencing.
to better understand their uniquely high levels of genetic diversity among endosymbionts.

**ACKNOWLEDGEMENTS**

This work was supported by funds from US National Science Foundation grant EF-0328363 to J. Werren, by funds from EU CSA-REGPOT 203590 – MicrobeGR and intramural funds of the University of Ioannina to K.B., by funds from US National Science Foundation grant IOS-0852344 and US National Institutes of Health grant RO1-GM085163-01 to S.B. 


**REFERENCES**


Wolbachia comparative genomics


candida (Hexapoda, Collembola) and its implications for wolbachial taxonomy. FEMS Microbiol Lett 180, 279–286.


Edited by: A. Holmes