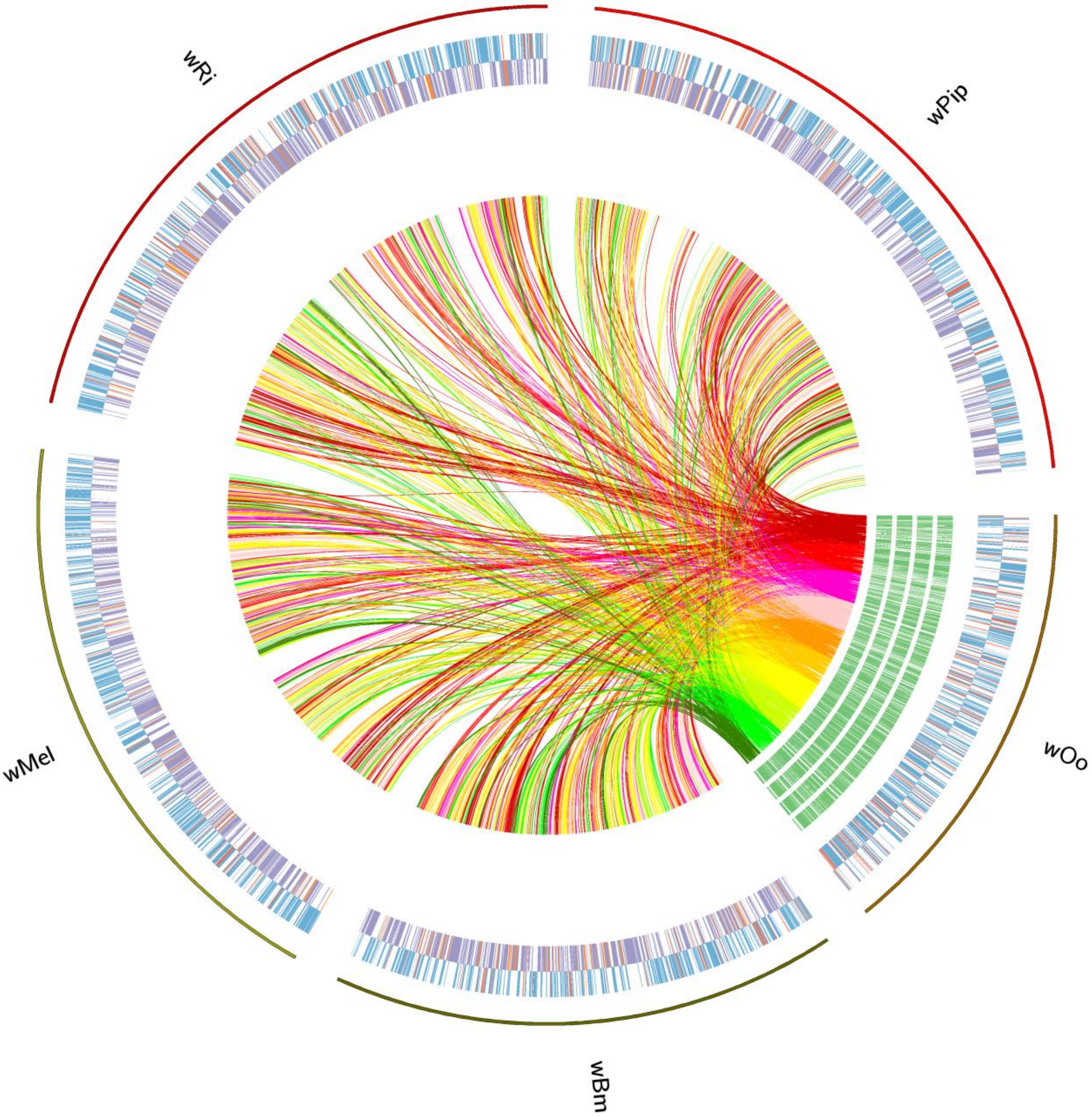
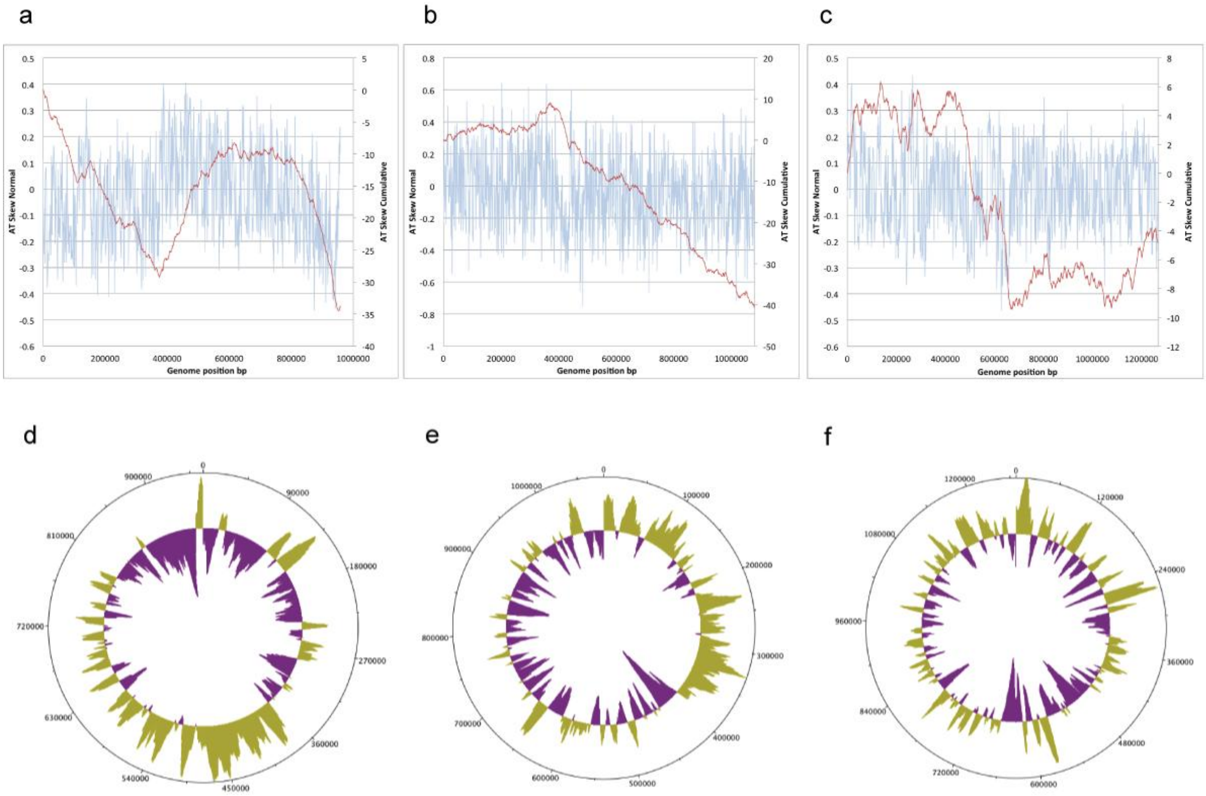
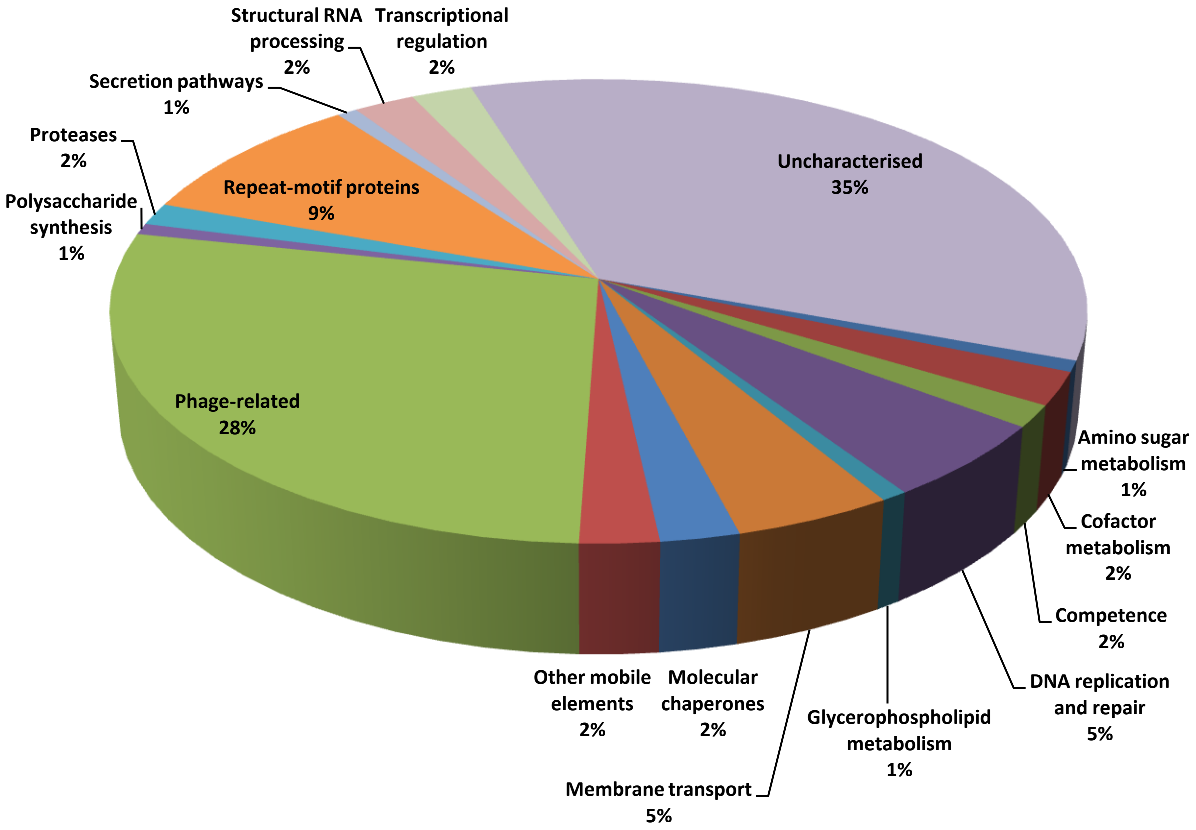
Table of Contents



Supplemental Fig. S1 Lack of macrosynteny between *Wolbachia* genomes. The syntenic regions between *w*Oo and the other completed *Wolbachia* genomes are shown by 10 uniquely-coloured, 100-kb windows (central zone). The first (outer) arc represents the linearized *Wolbachia* chromosome (to scale); the second and third arcs show the genes on the positive and negative strands, respectively (pseudogenes in red); and arcs 4 - 7 (in green) indicate the presence of *w*Oo genes in the other *Wolbachia* genomes (based on homology) in the order *w*Bm, *w*Mel, *w*Ri and *w*Pip.

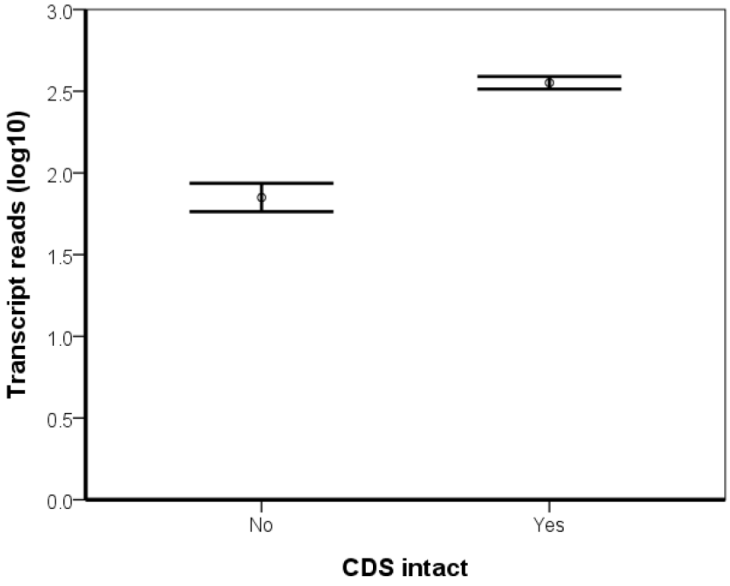


Supplemental Fig. S2 GC skew in *Wolbachia* genomes. Linear representation of the GC skew [(G-C)/(G+C)] using a 1,000-base window (a, b, c) and circular representation using a 10,000 base-window (d, e, f) for *w*Oo (a, d), *w*Bm (b, e) and *w*Mel (c, f). The *w*Oo and *w*Bm chromosomes exhibit a GC skew pattern or strand asymmetry that is typical of obligate mutualists, whereas this is not apparent for *w*Mel. Accordingly, in contrast with insect *Wolbachia*, the strains from filariae are entirely dependent on vertical transmission and almost certainly do not encounter other *Wolbachia* variants or symbiont species in their intracellular niche (Brownlie et al. 2007). Moreover, we observed a remarkable paucity of DNA repair mechanisms in *w*Oo, and analysis of strand compositional bias in the genomes of other obligate intracellular bacteria has indicated that the diminution of these pathways is strongly associated with the magnitude of GC skew (Klasson and Andersson 2006).

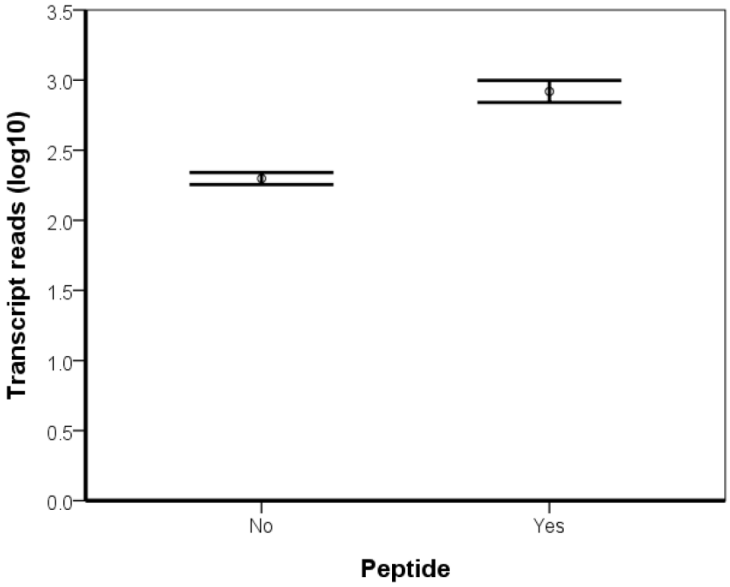


Supplemental Fig. S3 Orthologues uniquely conserved between insect *Wolbachia* strains. Orthologous cluster analysis identified 130 genes that are shared between the genomes of *w*Mel, *w*Ri and *w*Pip, but which are absent from those of *w*Bm and *w*Oo. These genes were manually classified into functional categories on the basis of conserved domain features obtained from the National Center for Biotechnology Information.

The genomes from the arthropod strains exhibit remarkable plasticity, in which ongoing recombination, gene duplication and bacteriophage infection have produced a mosaic structure (Klasson et al. 2008; Klasson et al. 2009; Wu et al. 2004). Accordingly, the 130 insect *Wolbachia* orthologues are dominated by phage-related structural genes and repeat-motif protein genes, together with an enlarged set of proteins involved in DNA replication, repair and competence that may facilitate recombination events. Among the repeat-motif proteins that have been the subject of numerous studies focused on potential interactions with the eukaryotic host (Papafotiou et al. 2011; Sinkins et al. 2005), the insect strains share 11 ankyrin-repeat (ANK) orthologues that are not found in the filarial symbionts. In contrast, *w*Oo exhibits a remarkable paucity of ANKs, retaining only two from the conserved repertoire of six that are present both in *w*Bm and the insect *Wolbachia*. These could represent either the minimum complement of ANKs required for the maintenance of symbiosis, or merely a remnant remaining from an ongoing process of gene degradation. The insect strains exhibit few uniquely conserved metabolic genes, despite the extensive differences in life history between the arthropod and nematode hosts. However, a haem oxygenase-like orthologue is present in all insect *Wolbachia* genomes, and in many bacteria this functions to scavenge iron from host-derived haem (Frankenberg-Dinkel 2004). In the symbiotic context, this enzyme could play a key role in the regulation of iron levels in insect hosts by recycling haem, as iron is usually limiting in the natural diet of *Drosophila*, and *Wolbachia* has been demonstrated to increase the fitness of this host by maintaining iron homeostasis (Brownlie et al. 2009). In addition, the insect *Wolbachia* genomes contain two enzymes involved in thiamine metabolism that are not present in *w*Bm and *w*Oo. Thiamine provisioning could be particularly important for blood-feeding arthropod hosts, as bedbugs are known to be dependent on their *Wolbachia* symbiont for B-vitamins during nymphal development (Hosokawa et al. 2010).



Supplemental Fig. S4 Quantitative mRNA expression from intact and pseudogenized genes. Error bars represent the mean ± 95% confidence interval for $\log_{10}$-transformed transcript reads. The difference in expression levels is statistically significant (independent samples *t*-test, equal variances not assumed, $p < 0.001$; data were analysed in PASW Statistics 18).



Supplemental Fig. S5 Quantitative mRNA expression from genes with or without peptide evidence of protein expression. Error bars represent the mean ± 95% confidence interval for $\log_{10}$-transformed transcript reads. The difference in expression levels is statistically significant (independent samples *t*-test, equal variances not assumed, $p < 0.001$; data were analysed in PASW Statistics 18).

Supplemental Table S1

Comparative genomic features of *Wolbachia*[a]

	***w*Oo**	***w*Bm**	***w*Mel**	***w*Ri**	***w*Pip**
Genome size (bp)	957,990	1,080,084	1,267,782	1,445,873	1,482,455
G+C content (%)	32.1	34.2	35.2	35.2	34.2
Predicted protein-coding genes[b]	842 (646; 196)	903 (805; 98)	1,269 (1,195; 74)	1,264 (1,150; 114)	1,385 (1,275; 110)
tRNA genes[b]	34 (33; 1)[c]	34	34	34	34
rRNA genes	3[c]	3	3	3	3
sRNA genes[d]	4[c]	4	5	5	5
Gene density (intact protein-coding genes per Kb)	0.68	0.75	0.94	0.80	0.86
Coding density (protein-coding nucleotides, %)	66.5	67.0	80.1	78.1	81.2

[a] Genome statistics for *w*Bm, *w*Mel, *w*Ri and *w*Pip were based on features defined by the current RefSeq record.

[b] (Genes with an intact CDS, putative pseudogenes).

[c] Full details are provided in Supplemental Table S2.

[d] The repertoire of sRNA genes in each genome was independently confirmed by BLAST analysis in the current study (group II introns are not included).

Supplemental Table S2

Structural RNA genes in the *w*Oo genome

RNA class	**RNA species**[a]
Ribosomal RNAs	23S rRNA
	16S rRNA
	5S rRNA
Transfer RNAs	Ala-TGC
	Gly-GCC
	Gly-TCC
	Pro-TGG
	Thr-GGT
	Thr-TGT
	Val-GAC
	Val-TAC
	Ser-GGA
	Ser-TGA
	Ser-GCT
	Arg-ACG
	Arg-CCG
	Arg-CCT
	Arg-TCT
	Leu-GAG[b]
	Leu-CAG
	Leu-TAG
	Leu-CAA
	Leu-TAA
	Phe-GAA
	Asn-GTT
	Lys-TTT
	Asp-GTC
	Glu-TTC
	His-GTG
	Gln-TTG
	Ile-GAT
	Met-CAT[c]
	Tyr-GTA
	Cys-GCA
	Trp-CCA
Small RNAs	Ribonuclease P
	Transfer-messenger RNA
	6S RNA
	4.5S RNA

[a] Each gene is present in a single copy unless otherwise stated.

[b] Predicted pseudogene.

[c] Present in three copies.

Supplemental Table S3

Insertion sequences identified in the *w*Oo genome

Genomic coordinates	IS name	IS family	Size (bp)	Orientation	Structure
130759 - 131474	ISWpi1	IS5	716	+	partial and degraded
414499 - 415386	ISWpi2	IS481	888	-	partial and degraded
503928 - 504374	ISWpi1	IS5	447	-	partial and degraded
660503 - 661444	ISWpi16	IS982	942	-	partial and degraded
869686 - 870660	ISWosp5	IS4 (IS231)	975	-	partial and degraded
950014 - 950678	ISWpi1	IS5	665	-	partial and degraded

Supplemental Table S4

Transposable element content (IS elements and group II introns) in fully sequenced *Wolbachia* genomes

	***Wolbachia* strain**				
	***w*Oo**	***w*Bm**	***w*Mel**	***w*Ri**	***w*Pip**
Supergroup	C	D	A	A	B
Lifestyle	Mutualist		Reproductive parasite		
Genome size (kb)	958	1,080	1,268	1,446	1,482
IS copy number	6	52[a,b]	105[b]	171[b]	170[b]
Genome coverage (%)	0.5	2.6	6.1	11.0	8.4
Group II intron copy number	0	0[c]	16[c]	14[c]	6[c]
Genome coverage (%)	0	0	1.9	1.8	0.8

(IS) insertion sequence.
[a] Reported by Cordaux 2009.
[b] Reported by Cerveau et al. 2011.
[c] Reported by Leclercq et al. 2011.

The IS elements in the wOo genome belong to four different families that have been reported previously from other *Wolbachia* genomes including ISWpi1 (3 copies in *w*Oo), an IS group which has recently experienced intense transpositional activity in arthropod *Wolbachia* strains (Cordaux et al. 2008). However, all six *w*Oo IS sequences are highly degraded, containing transposase genes that are pseudogenized and clearly non-functional. We conclude that there is apparently no ongoing transposable element activity in *w*Oo, which as a feature shared in common with the *w*Bm genome, suggests distinct evolutionary trajectories between the filarial and arthropod strains. Nevertheless, contrary to *w*Bm [where IS still occupy 2.6% of the genome (Cordaux 2009)], gene conversion does not seem to impact IS sequence evolution in *w*Oo, based on the high nucleotide divergence (~18%) recorded between *w*Oo ISWpi1 copies. Overall, *w*Oo appears to be at a very advanced stage of elimination of its transposable elements, which is fully consistent with the dramatically reduced genome size of this strain.

Supplemental Table S5

Enzymes involved in the biosynthesis of lipid II		
Escherichia coli	***w*Bm**[a]	***w*Oo**
MurA	wBm0740	wOo03240
MurB	wBm0778	wOo06610
MurC	wBm0118	wOo04220
MurD	wBm0508	wOo01420
MurE	wBm0492	wOo08850
MurF	wBm0238	wOo07780
MurG	wBm0557	wOo09500
MraY	wBm0643	wOo03800
Ddl	wBm0570	wOo01490
Alr	-	-
MurI	-	-

[a] Reported by Foster et al. 2005 and Henrichfreise et al. 2009.

The biosynthetic capability of *Wolbachia* for peptidoglycan assembly (or at least production of peptidoglycan precursors) is not fully understood, but is of considerable interest for fundamental prokaryotic cell biology, innate immunity, and the prediction of antibiotic efficacy. The pseudogenization of *argD* in *w*Oo may be significant in this context because in other *Wolbachia* genomes, *argD* is located within a truncated lysine biosynthesis pathway that generates the only amino-acid that can be produced by these strains, *meso*-diaminopimelate (DAP) (Foster et al. 2005). In most Gram-negative bacteria, *meso*-DAP occupies the third position in the pentapeptide moiety of lipid II and forms part of the ligand for the mammalian innate pattern recognition receptor, nucleotide-binding oligomerisation domain-containing protein 1 (Girardin et al. 2003). In the order *Chlamydiales*, *meso*-DAP is synthesized by an alternative pathway that utilises a L,L-DAP aminotransferase (McCoy et al. 2006); however, we found no evidence for this gene in the *w*Oo genome. Thus, the lipid II pentapeptide in *w*Oo may exhibit two aberrant features: a putative absence of D-amino-acid residues [in common with other *Wolbachia* strains (Foster et al. 2005)], and the substitution of *meso*-DAP with lysine or another amino-acid (obtained from the worm host); a structure more typical of some Gram-positive organisms (Schleifer and Kandler 1972). Genomes from all *Wolbachia* strains encode the full complement of genes required for synthesis of the peptidoglycan intermediate, lipid II. However, although *Wolbachia* contains genes for penicillin-binding protein 2 and D-alanyl-D-alanine carboxypeptidase, which could cross-link the peptide moieties of individual peptidoglycan building-blocks, it lacks genes for the transglycosylases required for full polymerisation of the glycan backbone in the cell wall (Foster et al. 2005; Henrichfreise et al. 2009). Moreover, alanine and glutamate racemases, responsible for the stereochemical conversion of these amino-acids, are also absent from *Wolbachia* genomes, suggesting that the pentapeptide component of lipid II may contain L-isomers of alanine and glutamate instead of the usual D-isomers (Foster et al. 2005). Whereas the insect *Wolbachia* genomes contain an *N*-acetylmuramoyl-L-alanine amidase that may be involved in terminal processing of a peptidoglycan macromolecule, and *w*Bm has an orthologue for this gene (wBm0682), no such orthologue is present in *w*Oo. The production of lipid II has been biochemically confirmed in a *Wolbachia* strain (*w*AlbB) from the mosquito *Aedes albopictus*; further, it has been proposed that *Wolbachia* has retained the capability to synthesise lipid II because it is essential for cell division, while assembly of fully polymerised peptidoglycan has been dispensed with due to the redundancy of the cell wall in an intracellular niche (Henrichfreise et al. 2009).

Supplemental Table S6 [see file "Supp Table S6.xls"]: The complete genome and transcriptome of *Wolbachia* strain *w*Oo, ranked by transcript abundance for protein-coding genes.

Note: *FIGfam categories (column AM) have been expanded using additional information obtained from the Kyoto Encyclopaedia of Genes and Genomes and conserved domains from the National Center for Biotechnology Information (NCBI). These manually-curated entries are highlighted in yellow.*

In free-living bacteria, transcripts for elongation factor (EF)-Tu account for ~10% of total RNA and usually represent the most abundant mRNA species (Fiola et al. 2006). Although proportionally this was not the case for the *w*Oo transcriptome, EF-TU mRNA from locus wOo10510 was still the single most abundant mRNA transcript. This protein is a GTPase that delivers aminoacyl-tRNA to the ribosome; following codon recognition, hydrolysis of GTP leads to dissociation of EF-Tu and entry of the cognate tRNA into the ribosomal A-site (Berisio et al. 2010). In addition, during *trans*-translation, EF-Tu binds to tmRNA (moderately expressed in *w*Oo) and facilitates the rescue of ribosomes that have stalled during the translation of aberrant transcripts (Hayes and Keiler 2010). In common with most other Gram-negative bacteria, *Wolbachia* has two distinct genes for EF-Tu (Weijland et al. 1992), which are 95% similar at the amino-acid level. Although expression from the second paralogue (wOo01560) was > 60% lower than for wOo10510 in the *w*Oo transcriptome, wOo01560 was also ranked within the top 10 most highly expressed genes.

Of the remaining highly abundant transcripts ranked within the top 20, five (*groEL*, *groES*, *clpB*, *dnaK* and *lon*) are chaperones or proteases associated with the heat-shock response. In free-living bacteria, this response is regulated by the sigma factor RpoH (also highly expressed in *w*Oo), and is characterized by a coordinated series of events that limits the accumulation of misfolded or aggregated proteins (Sabate et al. 2010). An ATPase, DnaK, mediates correct folding of nascent polypeptides as they emerge from the ribosome, and can repair damaged proteins in concert with the co-chaperone DnaJ and the nucleotide-release cofactor GrpE (both expressed at moderate levels in *w*Oo). However, proteins that are functionally dependent on a second phase of conformational rearrangement are transferred to the central cavity of the ELS complex, formed from the chaperonin GroEL and the co-chaperonin, GroES. Here, in another ATP-dependent process, the native conformation of the substrate protein is thermodynamically favoured and the spontaneous aggregation of larger proteins is inhibited. Under stress conditions where the protein quality-control system becomes exhausted, aggregates of misfolded proteins can be resolubilised by the coordinated action of DnaK and DnaJ with a 'disaggregase', ClpB (Tyedmers et al. 2010; Winkler et al. 2010). Finally, protein aggregates that are resistant to refolding are digested by ATP-dependent proteases, such as Lon (Garcia-Fruitos et al. 2007; Huang et al. 2001; Tomoyasu et al. 2001; Vera et al. 2005). Since non-silent mutations in coding sequences are a key process promoting protein misfolding and aggregation (Tyedmers et al. 2010), the loss of DNA repair mechanisms in *w*Oo is probably the principal driver of the constitutive heat-shock response. Indeed, an absence of DNA repair machinery has been shown experimentally to drive a reduction in genomic GC content (Lind and Andersson 2008) and other bacteria with highly degraded, AT-rich genomes are predicted to encode proteins with low thermodynamic stability (van Ham et al. 2003). Moreover, the relatively high level of expression from pseudogenes that we observed in *w*Oo could result in the production of truncated, non-functional peptides, as has been suggested for *Mycobacterium leprae* (Williams et al. 2009). Thus, the abundance of transcripts for the excinuclease UvrA in the *w*Oo transcriptome may represent a compensatory mechanism to control the mutation rate.

Other highly expressed genes in the *w*Oo transcriptome with defined roles in translation included two ribosomal proteins and a 2-methylthioadenine synthetase, MiaB. The latter enzyme catalyses the synthesis of 2-methylthio-*N*-6-isopentenyladenosine, a modified nucleoside found in certain tRNA transcripts (Hernandez et al. 2007). In addition, the abundance of transcripts for CspA suggests that this cold-shock protein may function in the global regulation of transcription [via its RNA chaperone activity, which destabilises RNA secondary structure (Phadtare 2004)] and translation [by binding to regions of mRNA required for initiation of translation (Horn et al. 2007)]. In contrast to the dominance of the translation machinery, we found that metabolic enzymes were generally expressed at lower levels in the *w*Oo transcriptome. Notable exceptions were adenylosuccinate synthetase (*purA*), which plays a critical role in *de novo* synthesis of purines; and the catalytic beta-subunit of F_oF_1 ATP synthase, the final complex of the respiratory chain.

Remarkably, four highly abundant transcripts from hypothetical genes were present in the *w*Oo transcriptome and two of these had no orthologues in the *w*Bm genome, but were represented in either *w*Pip alone (wOo07860) or in all of the completed insect *Wolbachia* genomes (wOo05320). Both of these genes were predicted to encode transmembrane proteins with signal peptides (Supplemental Table S14). In contrast, orthologues of wOo05220 were detected in all of the completed genomes except that of *w*Pip, whereas wOo06340 (a putatively cytoplasmic protein; Supplemental Table S14) was present in all sequenced strains.

Supplemental Table S7

Predicted essential genes absent from a *Wolbachia* genome

			Essentiality ranking[a]	
Locus tag in *w*Oo	**Annotation**	**Status**	**Multiple hit score**	**Gene conservation score**
wOo01260	Fumarate hydratase	No orthologue in *w*Ri	85	28
wOo03750	Recombinase A	Pseudogenized in *w*Oo	104	62
wOo03880	Glucosamine 6-phosphate synthetase	Pseudogenized in *w*Oo	35	475

(NA) not applicable.

[a] Data from Holman et al. 2009.

It is apparent that very few genes ranked within the top 100 for essentiality by Holman et al. 2009 are not present in all of the sequenced *Wolbachia* genomes (including that of *w*Oo), whether this was calculated using the Multiple Hit Score (based on the strength of alignments to known essential genes in other bacteria) or the Gene Conservation Score (based on the degree of phyletic conservation observed across the *Rickettsiales*). However, the three exceptions (fumarate hydratase*,* RecA, and glucosamine 6-phosphate synthetase) did not allow a meaningful comparison of the accuracy of the two scoring systems. The genes exhibiting the highest essentiality scores with either method are dominated by enzymes mediating DNA replication and transcription; elongation factors; aminoacyl tRNA synthetases; and enzymes involved in gluconeogenesis, the tricarboxylic acid cycle and amino-acid catabolism.

Supplemental Table S8

Type-IV secretion system genes in the *w*Oo genome		
***vir*B3 – B6 operon**	***vir*B8 – D4 operon**	**Extra-operonic**
wOo01290 VirB3	wOo07230 VirD4	wOo02080 VirB9
wOo01300 VirB4	wOo07240 VirB11	wOo02790 VirB4
wOo01310 VirB6a	wOo07250 VirB10	wOo03820 VirB8
wOo01320 VirB6b	wOo07260 VirB9	
wOo01340 VirB6c	wOo07270 VirB8	
wOo01350 VirB6d		

Supplemental Table S9

Additional genes differentially expressed between somatic and germline host tissue in *Wolbachia* strain *w*Oo

Tissue exhibiting upregulation	Locus tag	Annotation	Primary function	Fold-change
Soma	wOo05480	Ribonuclease P	Translation and RNA metabolism	6.20
	wOo06270	Hypothetical protein	Unknown	4.06
	wOo03560	Hypothetical protein	Unknown	2.05
Gonad	wOo09290	30S ribosomal protein S21	Translation	7.05
	wOo00370	Polypeptide deformylase	Translation	4.02
	wOo00360	Uracil-DNA glycosylase	DNA metabolism	3.20
	wOo07120	Preprotein translocase subunit SecF	Membrane transport	2.85
	wOo07540	6S RNA	RNA metabolism	2.71
	wOo00570	SAM-dependent methyltransferase	Translation	2.61
	wOo09590	Cardiolipin synthase[a]	DNA metabolism	2.50
	wOo08970	50S ribosomal protein L21	Translation	2.50
	wOo08520	F_oF_1-type ATP synthase, delta subunit[b]	Respiration	2.32
	wOo01610	50S ribosomal protein L1	Translation	2.23
	wOo01560	Elongation factor Tu	Translation	2.22
	wOo03050	Signal peptidase I	Membrane transport	2.21
	wOo10510	Elongation factor Tu	Translation	2.10
	wOo00910	5S rRNA	Translation	2.04
	wOo00090	Nucleoid DNA-binding protein	DNA metabolism	2.03
	wOo01170	Hypothetical protein	Unknown	1.91

(SAM) S-adenosyl-L-methionine.

[a] Predicted secreted protein (contains signal peptide) with conserved domain, “catalytic domain of EDTA-resistant nuclease Nuc from *Salmonella typhimurium* and similar proteins” (information from NCBI).

[b] This subunit can inhibit ATPase activity in *Micrococcus luteus* (Gruber et al. 1994), and thus the reduced transcript levels observed in somatic tissue could reflect regulatory uncoupling of the holoenzyme.

Several transcripts for hypothetical proteins were differentially expressed between the two tissue sites. These included the highly abundant mRNA for a hypothetical transmembrane protein not found in *w*Bm (wOo07860; Supplemental Table S14), which was upregulated 2.2-fold in soma, and a transcript for a protein containing a TrbC/VirB2 conserved domain (FC in soma, 3.1). The TrbC/VirB2 proteins are pilin subunits of the prototypic conjugative T4SS of *Agrobacterium tumefaciens* (Christie et al. 2005); however, pilus-like structures have never been observed in *Wolbachia*. Nonetheless, a close relative of *Wolbachia* that also lacks pili, *Anaplasma marginale*, also expresses VirB2 homologues that have been localised to the outer membrane (Sutten et al. 2010). Hence, it has been proposed that these may form surface appendages (not true pili), as has been demonstrated for VirB2 in *Helicobacter pylori* (Andrzejewska et al. 2006). Immunoelectron microscopic studies will be required to investigate whether these structures are present on the pleiomorphic forms of *Wolbachia* found in the hypodermal cords. We also sought to determine the tissue-specific expression of ANKs in *w*Oo, since data from arthropod hosts indicated that some of these proteins were differentially regulated between male and female gonads (*Wolbachia* are present in both in many insects), whereas no differences were apparent between male and female soma. Of the two *w*Oo ANKs, only wOo05150 showed > 2-fold upregulation in germline tissue and this difference, at 8% FDR, narrowly failed our significance cut-off. This observation supports the concept that ANKs are of greater importance in the interactions of *Wolbachia* with arthropod hosts than is the case for nematode hosts.

Supplemental Table S10

Gene models in the *w*Oo genome with peptide evidence[a]

Locus tag	Annotation	Coverage (%)	Unique peptides	PSM	Length (amino-acids)	MW (kDa)	pI
wOo08310	Cysteine desulfurase	13.51	4	4	444	49.8	8.69
wOo08860	Cystathionine beta-lyase	12.86	4	6	420	47.5	8.47
wOo09330	Hypothetical protein	24.51	4	4	204	23.5	7.02
wOo10240	DNA-directed RNA polymerase, subunit alpha	12.64	4	8	356	39.2	6.58
wOo00120	Outer membrane protein	11.17	3	4	403	45.1	9.13
wOo00510	F_oF_1 ATP synthase, subunit B	24.68	3	3	158	18.4	9.74
wOo01560	Elongation factor Tu	28.21	3	10	390	43.0	6.35
wOo03640	Outer membrane protein protective antigen OMA87	4.26	3	3	775	87.9	8.94
wOo06290	Membrane protease subunit stomatin/prohibitin-like protein	12.50	3	5	344	39.3	8.32
wOo06310	Trypsin-like serine protease	10.57	3	8	492	54.0	8.44
wOo06800	Malic enzyme	8.05	3	3	447	49.3	5.82
wOo07820	Polynucleotide phosphorylase/polyadenylase	3.84	3	5	755	84.1	6.76
wOo08210	Transcription termination factor Rho	6.37	3	3	471	52.4	8.57
wOo08520	F_oF_1 ATP synthase, delta subunit	19.68	3	3	188	21.6	9.73
wOo09770	Fructose 1,6-bisphosphatase II	10.65	3	5	310	33.4	6.43
wOo01240	AICAR transformylase/IMP cyclohydrolase PurH	3.30	2	2	515	57.5	7.68
wOo01630	50S ribosomal protein L7/L12	22.56	2	2	133	14.3	5.16
wOo01640	DNA-directed RNA polymerase, fusion of beta and beta' subunits RpoB/RpoC	0.92	2	3	2840	318.7	7.24
wOo02510	Thioredoxin trx	15.89	2	3	107	12.2	6.64
wOo02940	Molecular chaperone GrpE, heat shock protein	12.70	2	3	189	21.3	5.53
wOo03220	Prolyl-tRNA synthetase	4.24	2	2	425	48.5	8.57
wOo04800	Phosphatidylserine decarboxylase	11.30	2	3	230	26.1	8.29
wOo05380	Cell division protein FtsZ	7.05	2	3	397	42.8	5.45
wOo05460	Response regulator	10.00	2	2	250	28.1	6.80
wOo05630	Heat-shock protein 90	3.86	2	4	673	77.3	5.80
wOo05740	Actin-like ATPase involved in cell morphogenesis, MreB	6.98	2	2	358	38.4	5.85
wOo06000	Hypothetical protein	11.86	2	2	177	19.5	9.35

wOo06450	ATP-dependent protease, ATP-binding subunit	4.82	2	2	498	55.8	6.18
wOo06880	Dihydropteroate synthase, putative	4.86	2	2	432	48.8	8.10
wOo06980	NAD-specific glutamate dehydrogenase	1.71	2	2	1577	181.4	8.51
wOo07050	30S ribosomal protein S4	8.82	2	2	204	23.6	10.10
wOo07400	Histidyl-tRNA synthetase	8.96	2	2	413	46.9	8.41
wOo08200	Leucyl aminopeptidase	5.34	2	3	487	53.7	6.67
wOo08770	4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase	7.29	2	3	288	31.9	9.48
wOo09270	Succinyl-CoA synthetase, alpha subunit	8.56	2	3	292	30.6	6.61
wOo09300	Preprotein translocase, subunit SecB	22.75	2	3	167	18.7	5.07
wOo09430	ATP-dependent Lon protease	2.20	2	2	819	92.0	6.74
wOo09540	Alpha-ketoglutarate decarboxylase	2.37	2	2	886	100.5	7.47
wOo10320	30S ribosomal protein S5	15.29	2	2	170	18.4	10.42
wOo10350	30S ribosomal protein S8	18.32	2	2	131	15.0	10.04
wOo10750	Succinyl-diaminopimelate desuccinylase	6.23	2	2	401	44.2	8.05
wOo10760	Peroxiredoxin	11.56	2	3	199	22.4	5.21
wOo00090	Nucleoid DNA-binding protein	9.71	1	2	103	11.8	10.18
wOo00100	30S ribosomal protein S9	11.49	1	2	148	16.8	10.64
wOo00110	50S ribosomal protein L13	6.21	1	1	161	18.7	9.89
wOo00230	Inorganic pyrophosphatase	10.12	1	1	168	18.5	5.00
wOo00290	Ribonuclease D	3.26	1	1	399	46.5	8.46
wOo00500	F_oF_1 ATP synthase, subunit B	6.92	1	2	159	18.6	6.19
wOo01260	Fumarate hydratase	3.89	1	2	463	50.7	7.18
wOo01540	30S ribosomal protein S7	8.18	1	1	159	17.9	9.99
wOo01550	Elongation factor G	1.31	1	2	688	76.3	6.05
wOo01610	50S ribosomal protein L1	5.94	1	3	219	24.1	9.22
wOo01620	50S ribosomal protein L10	6.40	1	1	172	19.3	9.91
wOo01880	50S ribosomal protein L25	5.91	1	1	203	22.6	6.06
wOo02250	Threonyl-tRNA synthetase	2.21	1	1	633	72.4	6.67
wOo02740	Zn-dependent peptidase	4.02	1	1	423	48.1	7.03
wOo03040	Molecular chaperone DnaK	1.35	1	1	591	64.8	6.76
wOo03170	3-polyprenyl-4-hydroxybenzoate decarboxylase UbiX	9.62	1	1	208	22.9	9.17
wOo03240	UDP-N-acetylglucosamine enolpyruvyl transferase	2.82	1	1	426	46.1	7.66

wOo03250	FAD-dependent thymidylate synthase	3.48	1	1	287	34.1	9.13
wOo03500	Aspartate-semialdehyde dehydrogenase	4.09	1	2	342	38.5	7.94
wOo03560	Hypothetical protein	9.92	1	1	121	13.8	4.86
wOo03570	Isoprenoid biosynthesis protein	3.38	1	2	237	26.4	5.95
wOo03930	Phosphoribosylpyrophosphate synthetase	2.91	1	1	344	37.8	8.62
wOo04120	Phosphoglyceromutase	2.20	1	1	501	55.7	6.79
wOo04420	Major facilitator superfamily permease	3.94	1	1	406	45.1	8.69
wOo04520	Seryl-tRNA synthetase	3.70	1	1	432	49.9	8.66
wOo04980	3-oxoacyl-acyl-carrier-protein- synthase II	4.04	1	1	421	45.1	8.65
wOo05030	Dihydroorotate dehydrogenase 2	2.74	1	1	365	40.8	9.33
wOo05140	Succinate dehydrogenase, flavoprotein subunit	2.33	1	2	601	66.5	7.02
wOo05420	PASPAC domain-containing protein	3.17	1	1	378	44.6	9.20
wOo05590	5-aminolevulinate synthase	5.25	1	2	400	44.7	8.03
wOo05600	Hypothetical protein	4.96	1	1	262	28.3	7.03
wOo05780	Cytochrome c oxidase, subunit 2	6.69	1	1	254	28.9	8.35
wOo05880	Pyruvate dehydrogenase, subunit beta	7.23	1	1	332	36.1	5.57
wOo06390	Hypothetical protein	13.89	1	2	72	8.2	9.25
wOo06600	Porphobilinogen deaminase	4.45	1	1	292	32.7	8.81
wOo06820	Glutaredoxin-like protein	10.09	1	1	109	12.7	5.60
wOo06830	Cytochrome b subunit of the bc complex	2.44	1	1	410	47.6	9.10
wOo07010	50S ribosomal protein L28	8.87	1	1	124	14.4	11.30
wOo07240	Type-IV secretory pathway, VirB11 component	2.42	1	1	330	37.0	7.17
wOo07260	Type-IV secretory pathway, VirB9 component	3.36	1	1	268	31.3	4.94
wOo07440	Hypothetical protein	14.55	1	1	110	12.1	5.27
wOo07720	Malate dehydrogenase	3.80	1	1	316	34.0	6.80
wOo07920	dGTP triphosphohydrolase	4.26	1	2	399	46.2	6.89
wOo07980	Putative glycerol-3-phosphate acyltransferase PlsX	5.16	1	1	349	37.6	8.68
wOo08510	F_oF_1 ATP synthase, subunit alpha	5.85	1	8	513	56.6	7.47
wOo08680	Hypothetical protein	9.79	1	1	143	16.8	9.80
wOo09200	NADH ubiquinone oxidoreductase, 18 kDa subunit	12.12	1	1	99	11.5	10.13
wOo09240	Nitroreductase	5.88	1	1	187	21.4	6.34
wOo09440	ATP-dependent protease, ATP-binding subunit	3.06	1	1	425	46.7	6.48
wOo09960	Elongation factor Ts	5.15	1	1	291	32.5	6.62

wOo09970	30S ribosomal protein S2	5.67	1	1	282	31.4	8.98
wOo10000	Short-chain alcohol dehydrogenase family enzyme	4.02	1	1	249	27.5	5.16
wOo10220	ABC-type phosphate transport system, periplasmic component	4.14	1	1	338	38.5	8.91
wOo10300	Preprotein translocase, subunit SecY	5.19	1	1	366	41.1	9.41
wOo10330	50S ribosomal protein L18	10.48	1	2	124	14.6	10.52
wOo10370	50S ribosomal protein L5	4.52	1	2	177	20.1	9.32
wOo10420	50S ribosomal protein L16	9.49	1	1	137	15.3	10.78
wOo10440	50S ribosomal protein L22	7.76	1	1	116	13.0	10.32
wOo10470	50S ribosomal protein L23	8.42	1	1	95	11.2	10.33
wOo10480	50S ribosomal protein L4	5.39	1	1	204	23.1	9.92

(PSM) peptide-spectrum match

[a] The 20 most abundant proteins are listed separately in Supplemental Table S11.

Supplemental Table S11

Comparison of relative abundance of highly expressed proteins between *Wolbachia* strains *w*Oo and *w*Bm

Locus tag	**Annotation**	**No. unique peptides, *w*Oo**[a]	**Top 20 rank, *w*Bm**[b]	**No. unique peptides, *w*Bm**[b]
wOo06190	Chaperonin GroEL	45	Y	32
wOo08110	Outer membrane protein[c]	26	Y	11
wOo06990	ATP-binding subunit of Clp protease and DnaK/DnaJ chaperones	25	N	2
wOo07430	Outer surface protein WSP	17	Y	10
wOo05320	Hypothetical protein[c]	14	N	NO
wOo07860	Hypothetical protein[c]	11	N	NO
wOo05720	Outer membrane protein, PAL-like	9	Y	4
wOo09020	Molecular chaperone, DnaK	9	Y	12
wOo01710	Hypothetical protein[c]	8	N	1
wOo09280	Succinyl-CoA synthetase, beta subunit	8	N	1
wOo00850	F_oF_1 ATP synthase, subunit beta	7	Y	4
wOo05050	Fructose-bisphosphate aldolase	7	N	1
wOo06180	Co-chaperonin GroES	7	N	2
wOo02620	Dihydrolipoamide acyltransferase, E2 component	6	N	0
wOo06300	Membrane protease subunit stomatin/prohibitin-like protein	6	Y	4
wOo04430	Glycine/serine hydroxymethyltransferase	5	N	0
wOo10510	Elongation factor Tu	5	Y	6
wOo00040	Dihydrolipoamide dehydrogenase, E3 component	4	N	1
wOo04400	Aspartate aminotransferase	4	N	2
wOo08290	Bifunctional proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase	4	N	3

(WSP) *Wolbachia* surface protein; (PAL) peptidoglycan-associated lipoprotein; (NO) no orthologue present in *w*Bm genome.

[a] Data represent the top 20 most abundant proteins detected in the *w*Oo proteome.

[b] Data from Bennuru et al. 2011.

[c] Functional predictions for hypothetical proteins are provided in Supplemental Table S14.

The most striking divergence between *w*Oo and *w*Bm was in the proteomic representation of uncharacterised proteins, with those detected as very abundant transcripts in *w*Oo (wOo05320 and wOo07860) contributing 14 and 11 unique peptides, respectively. Whereas these proteins lack orthologues in *w*Bm, the most abundant uncharacterised proteins in *w*Bm (wBm0253 and wBm0771) are present in the *w*Oo genome but were not detected at the peptide level. Other marked differences between the proteomes included the presence of two highly abundant zinc-dependent peptidases in *w*Bm; one of these (wOo02740/wBm0221) was represented by just a single peptide in *w*Oo, while the other (wOo04600/wBm0419) was not detected. In common with *w*Oo, enzymes responsible for the generation of NTP, particularly ATP synthase, dominated the metabolic components of the *w*Bm proteome; but the greater abundance of individual enzymes from the *de novo* pyrimidine (carbamoylphosphate synthase) and nucleotide salvage (polyribonucleotide nucleotidyltransferase) pathways in *w*Bm suggest subtle differences in resource allocation between the two strains. Additionally, the more advanced state of genome reduction in *w*Oo compared with *w*Bm could underlie the increased profile of the heat-shock proteins, Clp protease and GroES, in the former.

Supplemental Table S12

Analysis of selection on *Wolbachia groEL*[a]

Sequence 1	**Sequence 2**	K_s	K_a	K_a/K_s[b]
*w*Mel	*w*Ri	0.0238	0.0008	0.034
*w*Mel	*w*Pip	0.5498	0.0429	0.078
*w*Mel	*w*Bm	0.5301	0.0361	0.068
*w*Mel	*w*Oo	0.5655	0.0462	0.082
*w*Ri	*w*Pip	0.5387	0.0437	0.081
*w*Ri	*w*Bm	0.5297	0.037	0.070
*w*Ri	*w*Oo	0.5735	0.0467	0.081
*w*Pip	*w*Bm	0.6111	0.0551	0.090
*w*Pip	*w*Oo	0.632	0.0646	0.102
*w*Bm	*w*Oo	0.4983	0.0479	0.096

[a] Selection tests on GroEL sequences were performed using DnaSP v. 5 (Librado and Rozas 2009).

[b] < 1, purifying selection; = 1, neutral evolution; > 1, positive selection.

Supplemental Table S13

Abundant proteins in the *w*Oo proteome with roles in amino-acid catabolism and sulphur metabolism

Locus tag	Annotation	Function	Reference
wOo02620	Dihydrolipoamide acyltransferase, E2 component	Control point of the TCA cycle; utilises amino-acid deamination products	Sonenshein 2007
wOo00040	Dihydrolipoamide dehydrogenase, E3 component		
wOo04400	Aspartate aminotransferase	Converts aspartate to glutamate	Tanous et al. 2005
wOo08290	Bifunctional proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase	Converts proline to glutamate	Tanner 2008
wOo08310	Cysteine desulfurase	Donates persulphide sulphur atoms to Fe-S clusters, thionucleosides, and lipoic acid	Hidese et al. 2011
wOo08860	Cystathionine beta-lyase	Converts methionine and cysteine to pyruvate	Lee et al. 2007

(TCA) tricarboxylic acid

Supplemental Table S14

***In silico* predictions for hypothetical protein-coding genes detected as abundant or differentially expressed transcripts, and/or as peptides**

	Subcellular localisation tool and prediction[a]					
Locus tag	**SOSUI$_{GramN}$**	**PSORTb**	**CELLO**	**ClubSub-P**	**(consensus)**	**Functional prediction**
wOo08110	EC	OM	OM	U	U	Porin domain[b,c]
wOo05320	OM	U	EC/OM/P	U	U	Signal peptide, transmembrane[c]
wOo07860	IM	U	P	U	U	Signal peptide, transmembrane[c]
wOo01710	EC	C	EC/OM	C	U	Putative modulator of DNA gyrase[b,c]
wOo06000	C	U	U	U	U	Transmembrane[c]
wOo07440	C	U	C	C	C	YbaB-like[b,c,d]
wOo03560	EC	U	C	U	U	Unknown
wOo05600	IM	EC	EC/OM/P	U	U	Transmembrane[c]
wOo08680	IM	C	IM	IM	IM	Predicted membrane protein[c]
wOo06270	IM	U	EC/OM/P	U	U	Unknown
wOo08460	IM	C	IM/C	IM	IM	TrbC/VirB2 family[e]
wOo09330	IM	C	C	IM	U	Copper chaperone SCO1/SenC[b,c]
wOo05220	IM	C	C	U	U	Unknown
wOo01170	EC	U	C/OM	C	U	Unknown
wOo06340	C	C	C	U	C	Unknown

(EC) extracellular; (OM) outer membrane; (IM) inner membrane; (P) periplasmic; (C) cytoplasmic; (U) unknown.

[a] These tools are described in Imai et al. 2008, Yu et al. 2010, Yu et al. 2006, and Paramasivam and Linke 2011 for SOSUI$_{GramN}$, PSORTb, CELLO and ClubSubP, respectively.

[b] Predicted by Phyre2.

[c] Predicted by InterProScan.

[d] DNA-binding protein (Cooley et al. 2009).

[e] Conserved domain information from NCBI.

Supplemental Table S15

Predicted lipoproteins in the genomes of *w*Oo and *w*Bm

		***In silico* prediction tool**[a]					**Consensus**[b]	
Locus tag	**Annotation**	**DOLOP**	**LIPO**	**LipoP**	**PRED-LIPO**	**LipPred**	***w*Oo**	***w*Bm**
wOo01320/wBm0795	Type-IV secretory pathway, VirB6 component	△	△			△	Y	N
wOo00410/wBm0606	Membrane protein related to metalloendopeptidase		△	△○	△○	△○	Y	Y
wOo07890[c]	Hypothetical protein		△	△	△		Y	-
wOo04200/wBm0117	Lipoprotein	△	△	△	△	△	Y	N
wOo01340/wBm0794	Type-IV secretory pathway, VirB6 component	△○	△○	△○			Y	Y
wOo01350/wBm0793	Type-IV secretory pathway, VirB6 component	△	△	△		△	Y	N
wOo05720/wBm0152	Outer membrane protein, PAL-like	○	△○	△○	○	△○	Y	Y
wOo09330/wBm0544	Hypothetical protein[d]			△○	△○	△○	Y	Y
wOo02540/wBm0268	Small protein A, tmRNA-binding	○	○			○	N	Y

(△) lipoprotein predicted in *w*Oo; (○) lipoprotein predicted in *w*Bm; (PAL) peptidoglycan-associated lipoprotein.

[a] These tools are described in Babu et al. 2006, Berven et al. 2006, Juncker et al. 2003, Bagos et al. 2008, and Taylor et al. 2006 for DOLOP, LIPO, LipoP, PRED-LIPO and LipPred, respectively.

[b] Agreement between three or more prediction tools was considered to be a consensus.

[c] No orthologue is present in *w*Bm.

[d] Functional prediction is provided in Supplemental Table S14.

Supplemental Methods

Nucleic Acid Extractions

For extraction of genomic DNA, approximately 150 male worms stored in PBS were washed twice in 1 ml 8 mM sodium hydroxide, and homogenised in DNAzol (Invitrogen) with a 1:1 ratio of 1 mm glass and 0.1 mm zirconia-silica beads in a Mini-Beadbeater (BioSpec Products). Genomic DNA was also obtained in a similar manner from ~20 mg of female worm gonads stored in RNA*later* (Sigma). The DNA was purified according to the manufacturer's instructions, solubilised in 8 mM sodium hydroxide, adjusted to pH 8.0 with HEPES free acid, and stored at 4°C. For extraction of total RNA, 100 – 150 male worms or ~25 mg female gonads in RNA*later* were homogenised in TRI Reagent (Sigma), using the Mini-Beadbeater as for DNA extraction. The RNA was purified according to the manufacturer's protocol, using Pellet Paint NF co-precipitant (Novagen) to maximise recovery. The RNA pellets were solubilised in nuclease-free water by incubation at 55°C for 10 min prior to storage at -80°C. Nucleic acids were quantified on a NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific).

Genome Sequence

The male worm genomic DNA was used to generate standard fragment and paired-end (8 kb insert) 454 DNA libraries using the GS Titanium Library Preparation Kit (Roche Applied Sciences) and sequenced on a Genome Sequencer FLX System (454 Roche Diagnostics). The data were assembled using Newbler (Roche Diagnostics) and after a preliminary analysis, the resulting contigs were filtered by BLAST to remove the non-*Wolbachia* sequences. The reads from these contigs were reassembled in Newbler and the filtered assembly (approximately 20× coverage) was edited and finished in GAP4, with reference to Sanger sequencing (GATC Biotech) of PCR products (generated using BIO-X-ACT Short Mix reagent, Bioline) spanning gaps in the assembly. The single contig representing the whole bacterial chromosome was then checked for integrity and quality using paired-end reads (100 bp ×2, 300 bp insert) obtained from a female gonad sample sequenced on a Genome Analyzer IIx platform (Illumina), which were mapped with Burrows-Wheeler Aligner. The resultant 50× coverage was used to correct 112 homopolymer length errors in the 454 assembly, producing a final coverage of 70×. Protein-coding genes were identified by PRODIGAL, while tRNA genes were located by tRNAscan-SE (Lowe and Eddy 1997). Putative functions were inferred using BLAST against the NCBI databases (Altschul et al. 1990) and InterProScan (Zdobnov and Apweiler 2001). Metabolic pathways were examined using the SEED (Overbeek et al. 2005) and Kyoto Encyclopaedia of Genes and Genomes databases (Kanehisa and Goto 2000), employing Artemis v. 13 to organize the data and facilitate annotation (Rutherford et al. 2000). Repeat identification was performed using MUMmer (Kurtz et al. 2004) and group II introns were queried as previously described (Leclercq et al. 2011), while IS elements were located using ISSaga (Varani et al. 2011) followed by manual verifications as detailed previously (Cerveau et al. 2011). Orthologues shared between *Wolbachia* were defined using ORTHOMCL (Li et al. 2003), whereas pseudogenes were classified as regions with homology to previously annotated bacterial genes, but that showed evidence of disruption by premature stop codons or frameshift mutations. In addition to the automated correction of the 454 assembly using Illumina genomic reads, putative pseudogenes were checked for data inconsistencies manually using both the Illumina dataset and SOLiD transcript reads (see below). Finally, mass spectrometric data from protein samples (see below) were searched against six-frame translations of highly transcribed pseudogene fragments and expressed intergenic regions, as identified by Cufflinks (Trapnell et al. 2010).

Synthesis and purification of cDNA

Contaminating genomic DNA in the RNA samples was digested using a TURBO DNA-*free* Kit (Ambion) according to the manufacturer's instructions. For first-strand cDNA synthesis, ~2.5 μg RNA was reverse-transcribed with random hexamer primers using a Moloney murine leukemia virus reverse-transcriptase (RT) kit (Bioline), as specified by the manufacturer. For each sample, RT- controls were prepared, in which a paired aliquot of RNA was incubated with an equivalent volume of nuclease-free water in place of RT. The first-strand reactions were performed at 40°C for 50 min and terminated at 70°C for 15 min. To confirm that first-strand cDNA was essentially free of genomic DNA contamination, transcripts of *ftsZ* and *rrsA* from *Wolbachia* were assayed by quantitative RT-PCR using SYBR Green I chemistry as previously described (Gilbert et al. 2005; Hansen et al. 2011). Copy numbers in the RT- controls were $\leq$ 0.1% of those in the equivalent cDNA samples. For second-strand synthesis, the complete first-strand reactions were incubated with 200 nM dNTP mix, 10 U *E. coli* DNA ligase, 40 U *E. coli* DNA polymerase I and 2 U *E. coli* RNase H in 1× Second-Strand Reaction Buffer (all supplied by Invitrogen) for 2 hr at 16°C. After addition of 10 U T4 DNA polymerase (Invitrogen), the incubation continued for 5 min at 16°C. The reactions were terminated by transfer onto ice with addition of EDTA (final concentration, 30 mM). The ds cDNA was purified by phenol-chloroform extraction, dissolved in nuclease-free water, and stored at -20°C.

Transcriptome sequencing and analysis

The ds cDNA samples (two biological replicates from male or female worms) were used to generate standard SOLiD fragment libraries and run as a multiplex on three SOLiD (v. 4; Applied Biosystems) slide quadrants in total. The resulting reads where mapped to the finished genome using Bowtie (Langmead et al. 2009). Transcript expression was analysed using Cufflinks (Trapnell et al. 2010) to find novel expressed regions in the genome, whereas differential expression (DE) was calculated using reads counts generated from bam2rpkm freeware (obtained from SourceForge) and analysed using edgeR (Bioconductor), with a negative binomial distribution model. For one male replicate, the number of tags mapping to the *w*Oo genome was insufficient for robust DE analysis, and thus this library was re-sequenced to greater depth on one-half of a SOLiD slide. For the identification of sRNA genes, non-coding regions exhibiting high levels of transcription were analysed using Rfam (Griffiths-Jones et al. 2003). To reduce the probability of including spurious DE, genes that exhibited < 1 count per million in four libraries were filtered from the analysis. Expression from 16S and 23S rRNA genes and tRNA genes was also removed to enable optimal normalisation of the data. In addition, small (< 150 bp) CDS predictions were excluded if they had no evidence of expression, whereas small CDS fragments showing expression within larger predicted pseudogenes were analysed at the level of the whole pseudogene.

Sample preparation for proteomic analysis

To provide *Wolbachia*-enriched samples for proteomic analysis, 600 mg of fresh adult female *O. ochengi* were processed using the FOCUS Mitochondria kit (G-Biosciences) according to the manufacturer's instructions for soft tissues. A protease inhibitor cocktail (FOCUS-Protease *Arrest*, G-Biosciences) was included during the initial homogenisation, and an optional secondary purification by gradient cushion was applied. Initial evaluation of the protocol showed that although bovine and filarial proteins continued to dominate the sample, the number of proteins identified as of endosymbiont origin was increased by > 3-fold compared with unfractionated worm preparations. The *Wolbachia*-enriched samples were mixed with an equal volume of reducing sample buffer, and the proteins were separated by PAGE on a 15% resolving gel, followed by post-electrophoretic visualisation using a Colloidal Blue Staining Kit (Invitrogen). The gel lane (loaded with 30 µg total protein) was cut into 30 slices for in-gel digestion with trypsin, and the peptides were cleaned with C18 Stage Tips (Proxeon) before pooling into five fractions for mass spectrometric (MS) analysis.

Liquid chromatography (LC)-MS and data analysis

Peptide mixtures were analyzed by on-line nanoflow LC using the nanoACQUITY-nLC system (Waters Corp.) coupled to an LTQ-Orbitrap Velos (Thermo Fisher Scientific) mass spectrometer equipped with the manufacturer's nanospray ion source. The analytical column (nanoACQUITY UPLC^{TM} BEH130 C18, 15 cm × 75 µm, 1.7 µm capillary column) was maintained at 35°C and a flow-rate of 300 nl/min. The gradient consisted of 3 - 40% acetonitrile in 0.1% formic acid for 90 min, followed by a ramp of 40 - 85% acetonitrile in 0.1% formic acid for 3 min. Full scan MS spectra (m/z range, 300 - 2000) were acquired by the Orbitrap at a resolution of 30,000. Analysis was performed in data-dependant mode. The top 20 most intense ions from MS1 scan (full MS) were selected for tandem MS by collision-induced dissociation and all product spectra were acquired in the LTQ ion trap. Ion trap and Orbitrap maximal injection times were set to 50 ms and 500 ms, respectively. Peptide identifications were established using the Mascot (v. 2.2.03, Matrix Science) search engine via the vendor-supplied Proteome Discoverer software (v. 1.2, Thermo Fisher Scientific). Tandem MS data were searched against a custom database that contained predicted proteomes from *w*Oo (complete genome; EMBL HE660029), the bovine genome (International Protein Index, v. 3.66), *Brugia malayi* (RefSeq genome; NCBI accession PRJNA27801) and *Onchocerca* EST libraries generated by the River Blindness Genome Project (Williams et al. 2002). Search parameters included a precursor mass tolerance of 10 ppm and fragment mass tolerance of 0.5 Da, while one missed tryptic cleavage was permitted. Carbamidomethylation was set as a fixed modification and oxidation (M) was included as a variable modification. A significance threshold of $p < 0.01$ was applied to the Mascot ion score and the false discovery rates were set at < 1%. Tentative identifications with only one unique peptide were manually validated considering mass accuracy, the assignment of major peaks, and the occurrence of uninterrupted y- or b-ion series of at least 3 consecutive amino acids. The number of unique peptides identified per protein was used as an estimate of abundance.

***In silico* analysis of protein-coding sequences**

To predict the subcellular localisation of hypothetical proteins from *w*Oo, four online tools were employed [SOSUI$_{\text{GramN}}$ (Imai et al. 2008), PSORTb (v 3.0.2; Yu et al. 2010), CELLO (Yu et al. 2006) and ClubSub-P (Paramasivam and Linke 2011)], and agreement between three or more programs was considered to be a consensus. Structural predictions of function for highly and/or differentially expressed hypothetical proteins from *w*Oo were performed using InterProScan (Zdobnov and Apweiler 2001) and Phyre2 (Soding 2005). Finally, the *w*Oo and *w*Bm genomes were screened for putative lipoproteins using five bioinformatic tools [DOLOP (Babu et al. 2006), LIPO (Berven et al. 2006), LipoP (Juncker et al. 2003), PRED-LIPO (Bagos et al. 2008) and LipPred (Taylor et al. 2006)].

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Andrzejewska J, Lee SK, Olbermann P, Lotzing N, Katzowitsch E, Linz B, Achtman M, Kado CI, Suerbaum S, Josenhans C. 2006. Characterization of the pilin ortholog of the *Helicobacter pylori* type IV *cag* pathogenicity apparatus, a surface-associated protein expressed during infection. *J Bacteriol* **188:** 5865-5877.

Babu MM, Priya ML, Selvan AT, Madera M, Gough J, Aravind L, Sankaran K. 2006. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol* **188:** 2761-2773.

Bagos PG, Tsirigos KD, Liakopoulos TD, Hamodrakas SJ. 2008. Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *J Proteome Res* **7:** 5082-5093.

Bennuru S, Meng Z, Ribeiro JM, Semnani RT, Ghedin E, Chan K, Lucas DA, Veenstra TD, Nutman TB. 2011. Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proc Natl Acad Sci U S A* **108:** 9649-9654.

Berisio R, Ruggiero A, Vitagliano L. 2010. Elongation factors EFIA and EF-Tu: their role in translation and beyond. *Israel Journal of Chemistry* **50:** 71-79.

Berven FS, Karlsen OA, Straume AH, Flikka K, Murrell JC, Fjellbirkeland A, Lillehaug JR, Eidhammer I, Jensen HB. 2006. Analysing the outer membrane subproteome of *Methylococcus capsulatus* (Bath) using proteomics and novel biocomputing tools. *Arch Microbiol* **184:** 362-377.

Brownlie JC, Adamski M, Slatko B, McGraw EA. 2007. Diversifying selection and host adaptation in two endosymbiont genomes. *BMC Evol Biol* **7:** 68.

Brownlie JC, Cass BN, Riegler M, Witsenburg JJ, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL. 2009. Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. *PLoS Pathog* **5:** e1000368.

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R. 2011. Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. *Genome Biol Evol* **3:** 1175-1186.

Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E. 2005. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* **59:** 451-485.

Cooley AE, Riley SP, Kral K, Miller MC, DeMoll E, Fried MG, Stevenson B. 2009. DNA-binding by *Haemophilus influenzae* and *Escherichia coli* YbaB, members of a widely-distributed bacterial protein family. *BMC Microbiol* **9:** 137.

Cordaux R. 2009. Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont. *Mol Biol Evol* **26:** 1679-1682.

Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol* **25:** 1889-1896.

Fiola K, Perreault JP, Cousineau B. 2006. Gene targeting in the Gram-positive bacterium *Lactococcus lactis*, using various delta ribozymes. *Appl Environ Microbiol* **72:** 869-879.

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, et al. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* **3:** e121.

Frankenberg-Dinkel N. 2004. Bacterial heme oxygenases. *Antioxidants & Redox Signaling* **6:** 825-834.

Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A. 2007. Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. *J Mol Biol* **374:** 195-205.

Gilbert J, Nfon CK, Makepeace BL, Njongmeta LM, Hastings IM, Pfarr KM, Renz A, Tanya VN, Trees AJ. 2005. Antibiotic chemotherapy of onchocerciasis: in a bovine model, killing of adult parasites requires a sustained depletion of endosymbiotic bacteria (*Wolbachia* species). *J Infect Dis* **192:** 1483-1493.

Girardin SE, Boneca IG, Carneiro LA, Antignac A, Jehanno M, Viala J, Tedin K, Taha MK, Labigne A, Zahringer U, et al. 2003. Nod1 detects a unique muropeptide from Gram-negative bacterial peptidoglycan. *Science* **300:** 1584-1587.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* **31:** 439-441.

Gruber G, Engelbrecht S, Junge W, Dose K, Nawroth T. 1994. Purification and characterization of the inhibitory subunit (delta) of the ATP-synthase from *Micrococcus luteus*. *FEBS Lett* **356:** 226-228.

Hansen RD, Trees AJ, Bah GS, Hetzel U, Martin C, Bain O, Tanya VN, Makepeace BL. 2011. A worm's best friend: recruitment of neutrophils by *Wolbachia* confounds eosinophil degranulation against the filarial nematode *Onchocerca ochengi*. *Proc R Soc B* **278:** 2293-2302.

Hayes CS, Keiler KC. 2010. Beyond ribosome rescue: tmRNA and co-translational processes. *FEBS Lett* **584:** 413-419.

Henrichfreise B, Schiefer A, Schneider T, Nzukou E, Poellinger C, Hoffmann TJ, Johnston KL, Moelleken K, Wiedemann I, Pfarr K, et al. 2009. Functional conservation of the lipid II biosynthesis pathway in the cell wall-less bacteria *Chlamydia* and *Wolbachia*: why is lipid II needed? *Mol Microbiol* **73:** 913-923.

Hernandez HL, Pierrel F, Elleingand E, Garcia-Serres R, Huynh BH, Johnson MK, Fontecave M, Atta M. 2007. MiaB, a bifunctional radical-S-adenosylmethionine enzyme involved in the thiolation and methylation of tRNA, contains two essential [4Fe-4S] clusters. *Biochemistry* **46:** 5140-5147.

Hidese R, Mihara H, Esaki N. 2011. Bacterial cysteine desulfurases: versatile key players in biosynthetic pathways of sulfur-containing biofactors. *Appl Microbiol Biotechnol* **91:** 47-61.

Holman AG, Davis PJ, Foster JM, Carlow CK, Kumar S. 2009. Computational prediction of essential genes in an unculturable endosymbiotic bacterium, *Wolbachia* of *Brugia malayi*. *BMC Microbiol* **9:** 243.

Horn G, Hofweber R, Kremer W, Kalbitzer HR. 2007. Structure and function of bacterial cold shock proteins. *Cell Mol Life Sci* **64:** 1457-1470.

Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* **107:** 769-774.

Huang HC, Sherman MY, Kandror O, Goldberg AL. 2001. The molecular chaperone DnaJ is required for the degradation of a soluble abnormal protein in *Escherichia coli*. *J Biol Chem* **276:** 3920-3928.

Imai K, Asakawa N, Tsuji T, Akazawa F, Ino A, Sonoyama M, Mitaku S. 2008. SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. *Bioinformation* **2:** 417-421.

Juncker AS, Willenbrock H, Von HG, Brunak S, Nielsen H, Krogh A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12:** 1652-1662.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28:** 27-30.

Klasson L, Andersson SG. 2006. Strong asymmetric mutation bias in endosymbiont genomes coincide with loss of genes for replication restart pathways. *Mol Biol Evol* **23:** 1031-1039.

Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, et al. 2008. Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. *Mol Biol Evol* **25:** 1877-1887.

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, et al. 2009. The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A* **106:** 5725-5730.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* **5:** R12.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Leclercq S, Giraud I, Cordaux R. 2011. Remarkable abundance and evolution of mobile group II introns in *Wolbachia* bacterial endosymbionts. *Mol Biol Evol* **28:** 685-697.

Lee WJ, Banavara DS, Hughes JE, Christiansen JK, Steele JL, Broadbent JR, Rankin SA. 2007. Role of cystathionine beta-lyase in catabolism of amino acids to sulfur volatiles by genetic variants of *Lactobacillus helveticus* CNRZ 32. *Appl Environ Microbiol* **73:** 3034-3039.

Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13:** 2178-2189.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci U S A* **105:** 17878-17883.

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

McCoy AJ, Adams NE, Hudson AO, Gilvarg C, Leustek T, Maurelli AT. 2006. L,L-diaminopimelate aminotransferase, a trans-kingdom enzyme shared by *Chlamydia* and plants for synthesis of diaminopimelate/lysine. *Proc Natl Acad Sci U S A* **103:** 17909-17914.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691-5702.

Papafotiou G, Oehler S, Savakis C, Bourtzis K. 2011. Regulation of *Wolbachia* ankyrin domain encoding genes in *Drosophila* gonads. *Res Microbiol* **162:** 764-772.

Paramasivam N, Linke D. 2011. ClubSub-P: Cluster-based subcellular localization prediction for Gram-negative bacteria and Archaea. *Front Microbiol* **2:** 218.

Phadtare S. 2004. Recent developments in bacterial cold-shock response. *Curr Issues Mol Biol* **6:** 125-136.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944-945.

Sabate R, de Groot NS, Ventura S. 2010. Protein folding and aggregation in bacteria. *Cellular and Molecular Life Sciences* **67:** 2695-2715.

Schleifer KH, Kandler O. 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36:** 407-477.

Sinkins SP, Walker T, Lynd AR, Steven AR, Makepeace BL, Godfray HC, Parkhill J. 2005. *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature* **436:** 257-260.

Soding J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21:** 951-960.

Sonenshein AL. 2007. Control of key metabolic intersections in *Bacillus subtilis*. *Nat Rev Microbiol* **5:** 917-927.

Sutten EL, Norimine J, Beare PA, Heinzen RA, Lopez JE, Morse K, Brayton KA, Gillespie JJ, Brown WC. 2010. *Anaplasma marginale* type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine. *Infect Immun* **78:** 1314-1325.

Tanner JJ. 2008. Structural biology of proline catabolism. *Amino Acids* **35:** 719-730.

Tanous C, Gori A, Rijnen L, Chambellon E, Yvon M. 2005. Pathways for alpha-ketoglutarate formation by *Lactococcus lactis* and their role in amino acid catabolism. *International Dairy Journal* **15:** 759-770.

Taylor PD, Toseland CP, Attwood TK, Flower DR. 2006. LIPPRED: A web server for accurate prediction of lipoprotein signal sequences and cleavage sites. *Bioinformation* **1:** 176-179.

Tomoyasu T, Mogk A, Langen H, Goloubinoff P, Bukau B. 2001. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol Microbiol* **40:** 397-413.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511-515.

Tyedmers J, Mogk A, Bukau B. 2010. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* **11:** 777-788.

van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, et al. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci U S A* **100:** 581-586.

Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* **12:** R30.

Vera A, Aris A, Carrio M, Gonzalez-Montalban N, Villaverde A. 2005. Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. *J Biotechnol* **119:** 163-171.

Weijland A, Harmark K, Cool RH, Anborgh PH, Parmeggiani A. 1992. Elongation factor Tu: a molecular switch in protein biosynthesis. *Mol Microbiol* **6:** 683-688.

Williams DL, Slayden RA, Amin A, Martinez AN, Pittman TL, Mira A, Mitra A, Nagaraja V, Morrison NE, Moraes M, et al. 2009. Implications of high level pseudogene transcription in *Mycobacterium leprae*. *BMC Genomics* **10:** 397.

Williams SA, Laney SJ, Lizotte-Waniewski M, Bierwert LA. 2002. The River Blindness Genome Project. *Trends in Parasitology* **18:** 86-90.

Winkler J, Seybert A, Konig L, Pruggnaller S, Haselmann U, Sourjik V, Weiss M, Frangakis AS, Mogk A, Bukau B. 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J* **29:** 910-923.

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, et al. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2:** E69.

Yu CS, Chen YC, Lu CH, Hwang JK. 2006. Prediction of protein subcellular localization. *Proteins* **64:** 643-651.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, et al. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26:** 1608-1615.

Zdobnov EM, Apweiler R. 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17:** 847-848.