

An Ancient Horizontal Gene Transfer between Mosquito and the Endosymbiotic Bacterium *Wolbachia pipientis*

Megan Woolfit, Iñaki Iturbe-Ormaetxe, Elizabeth A. McGraw, and Scott L. O'Neill

School of Integrative Biology, University of Queensland, Brisbane, Queensland, Australia

The extent and biological relevance of horizontal gene transfer (HGT) in eukaryotic evolution remain highly controversial. Recent studies have demonstrated frequent and large-scale HGT from endosymbiotic bacteria to their hosts, but the great majority of these transferred genes rapidly become nonfunctional in the recipient genome. Here, we investigate an ancient HGT between a host metazoan and an endosymbiotic bacterium, *Wolbachia pipientis*. The transferred gene has so far been found only in mosquitoes and *Wolbachia*. In mosquitoes, it is a member of a gene family encoding candidate receptors required for malaria sporozoite invasion of the mosquito salivary gland. The gene copy in *Wolbachia* has substantially diverged in sequence from the mosquito homolog, is evolving under purifying selection, and is expressed, suggesting that this gene is also functional in the bacterial genome. Several lines of evidence indicate that the gene may have been transferred from eukaryotic host to bacterial endosymbiont. Regardless of the direction of transfer, however, these results demonstrate that interdomain HGT may give rise to functional, persistent, and possibly evolutionarily significant new genes.

Introduction

Vertically transmitted endosymbiotic bacteria are an ideal model system for investigating horizontal gene transfer (HGT) between prokaryotes and eukaryotes. The transovarial route of transmission of these bacteria means that they are intimately associated with host germ tissue over long periods of evolutionary time, which is expected to enhance the frequency of HGT events. Biological interdependence of host and symbiont may also increase the likelihood that transferred genes are functionally relevant to the recipient and thus maintained in its genome. *Wolbachia pipientis* is an α -proteobacterial obligate endosymbiont of arthropods and nematodes that is estimated to infect at least 20% of insect species (Werren and Windsor 2000) and possibly as many as 66% (Hilgenboecker et al. 2008). Recent studies have demonstrated that genetic sequences, ranging in size from single genes to entire bacterial genomes, have been transferred from *Wolbachia* to many of their insect hosts (Kondo et al. 2002; Fenn et al. 2006; Dunning Hotopp et al. 2007). Given *Wolbachias* estimated prevalence, interdomain gene transfer may occur much more frequently than has previously been thought.

Most of the genes transferred between *Wolbachia* and their hosts are nonfunctional in the recipient genome, as these gene copies often contain stop codons, frameshifts, or retroelement insertions and/or are transcriptionally inactive (Dunning Hotopp et al. 2007; Nikoh et al. 2008). The transferred genes therefore do not typically fulfill the basic criteria for evolutionary significance: longevity and integration into the biology of the recipient taxon (Blaxter 2007). In contrast, the family of salivary gland surface (SGS) proteins previously characterized in *Aedes aegypti* and *Anopheles gambiae* meet both of these criteria.

SGS genes have no known nonmosquito eukaryotic homologs but share moderate sequence similarity with the hypothetical gene *WD0513* from *wMel*, the *Wolbachia* endosymbiont of *Drosophila melanogaster* (Wu et al.

2004). This has led to the suggestion that these genes arose as a result of an ancient HGT, probably from *Wolbachia* to mosquito (Arca et al. 2005; Korochkina et al. 2006). The proteins encoded by SGS genes play a role in insect–malaria interactions (Korochkina et al. 2006). After a female mosquito takes a malaria-infected blood meal, malarial sporozoites must invade the salivary glands before the parasite can be transmitted to a new host. This invasion is thought to require specific interactions between sporozoite ligands and mosquito salivary gland receptors (Brennan et al. 2000). *Aedes aegypti* aaSGS1 (encoded by *AAEL009993*), for example, localizes to the basal lamina of the salivary glands of adult female mosquitoes, and inoculation of mosquitoes with anti-aaSGS1 antibodies reduces sporozoite invasion by approximately 65% (Korochkina et al. 2006), indicating that sporozoites interact with aaSGS1 during salivary gland invasion. SGS homologs are also expressed in adult female salivary glands in *Aedes albopictus* (Arca et al. 2007), *An. gambiae* (Arca et al. 2005; Korochkina et al. 2006), and *Anopheles funestus* (Calvo et al. 2007), where they may perform similar functions.

Here, we use the complete genome sequences of the mosquitoes *Ae. aegypti*, *An. gambiae*, and *Culex quinquefasciatus*, and of *Wolbachia* strains infecting *D. melanogaster* (*wMel*) and *Cx. quinquefasciatus* (*wPip*), to confirm that the SGS genes have been horizontally transferred between these taxa. To determine the direction of HGT, we then consider data on the taxonomic distribution, phylogenetic relationships, gene length, GC content, synonymous codon usage, and intronic content of SGS genes. Taken together, we believe that the balance of evidence supports a direction of transfer from eukaryotic host to *Wolbachia* endosymbiont.

Materials and Methods

Sequence Data

To identify all SGS gene homologs in mosquitoes, we used the amino acid sequences of aaSGS1 (GenBank accession number AAV28546) and *WD0513* (GenBank NP_966293) as BLASTP queries against the predicted peptide databases of *Ae. aegypti* (gene build AegL1.1), *An. gambiae* (gene build AgamP3.4), and *Cx. quinquefasciatus* (gene build CpipJ1.1), downloaded from Vectorbase

Key words: endosymbiont, symbiosis, *Wolbachia pipientis*, *Aedes aegypti*, *Anopheles gambiae*, SGS.

E-mail: scott.oneill@uq.edu.au.

Mol. Biol. Evol. 26(2):367–374. 2009

doi:10.1093/molbev/msn253

Advance Access publication November 6, 2008

(<http://www.vectorbase.org>). To identify any unannotated homologs, we also used these two sequences as TBLASTN queries against the three assembled mosquito genome sequences, downloaded from Vectorbase. We extracted all hits from each search with *E*-values better than $1e-05$. To confirm that hits were more closely related to SGS genes than to members of other gene families, we used their translations as BLASTP queries against the nonredundant (NR) database and checked that their best hit was either another mosquito SGS gene or *WD0513*. All mosquito homologs are listed in supplementary table S1, Supplementary Material online.

During this analysis, we noticed that many of the mosquito SGS genes returned a match to *WD0512* in addition to *WD0513*. These genes are adjacent and overlapping (by 86 bp) in the *wMel* genome. We used *WD0512* as a TBLASTN query against the mosquito genomes and found that the first 300–500 bp of this gene was homologous to the 3' end (or sometimes the 3' untranslated region [UTR]) of each SGS gene. The 3'-UTR of *AAEL004181* was homologous to the full length of *WD0512*.

To identify homologs of *WD0513* in other *Wolbachia* strains, we used the amino acid sequence encoded by this gene as a BLASTP query against the *wMel* and *wBm* proteomes and as a TBLASTN query against the genome assemblies of the *Wolbachia* strains *wWil*, *wAna*, and *wSim* (all http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), *wOv* (www.sanger.ac.uk/Projects/Wolbachia/), and *wPip* (http://www.sanger.ac.uk/Projects/W_pipientis/). Only a single homolog was identified, in *wPip*. This gene, recently annotated as *WP1346* (Klasson et al. 2008), is homologous to almost the full length of *WD0513* and *WD0512*.

Phylogenetic Analysis

To identify possible outgroup sequences for our phylogenetic analysis, we used the amino acid sequences of *WD0513* and each of the mosquito SGS proteins as BLASTP or TBLASTN queries against the NR, genomic survey sequences (GSS), whole genome shotgun reads, high throughput genomic sequences, and environmental sequence databases on NCBI (<http://www.ncbi.nlm.nih.gov>). We chose as outgroups four proteins that were in the top 10 nonmosquito hits for all queries against the NR database (no closer outgroups were identified in any of the other databases): *Flavobacterium johnsoniae* YD-repeat-containing protein, GenBank accession number YP_001196062.1; *Bacteroides thetaiotaomicron* conserved hypothetical protein, NP_811839.1; *Bacteroides ovatus* hypothetical protein, EDO13775.1; and *Cellulophaga* sp. cell wall associated RhsD protein, ZP_01049542.1.

We excluded both of the unannotated SGS homologs we identified (supplementary table S1, Supplementary Material online) from the phylogenetic analysis, as only fragmentary gene sequence was available for each. We also excluded three of the annotated genes, *AAEL003694*, *CPIJ007816*, and *CPIJ004852* (although inclusion of these genes does not change our conclusions with regard to the direction of transfer; supplementary fig. S1, Supplementary Material online). *AAEL003694* is a very short, truncated gene (coding for 846 residues, in comparison with, e.g., aaSGS1 with 3,060 residues). *CPIJ007816* contains two tandem segments, each homologous to almost the full

length of *WD0513*. We believe that *CPIJ007816* probably represents two separate genes, but the true gene boundaries cannot be clearly determined. *CPIJ004852* is a highly divergent gene that is very difficult to align with confidence and whose phylogenetic position varies with the method of alignment and outgroup choice.

The remaining sequences were translated into amino acids and aligned using T-coffee (Notredame et al. 2000). Regions of good sequence conservation were observed along the full length of the ingroup genes. Only the most highly conserved domain could be confidently aligned between ingroup and outgroup genes, and so this region alone was used for the phylogenetic analysis. Ingroup topology remains the same if outgroups are not included and a full-length alignment of ingroup sequences is used (data not shown). The alignment was edited by hand in Se-AL (Rambaut 1996) and then trimmed using Gblocks (Castresana 2000) with a range of stringency settings. All settings gave very similar results after phylogenetic analysis; we report results of settings $-b3 = 15$ and $-b5 = a$, which allow the retention of sites at which some sequences have a gap and removes any regions of the alignment containing more than 15 contiguous poorly conserved sites. This produced an alignment of 454 aa.

We constructed Bayesian phylogenetic trees using MrBayes (Huelsenbeck and Ronquist 2001), performing five runs each of four chains, sampling every 100th iteration over 10,000,000 generations after a burn-in of 25,000 samples and allowing mixed models of amino acid substitution. Maximum likelihood trees were constructed with PHYML (Guindon and Gascuel 2003), using the Jones, Taylor, and Thornton model of substitution and a gamma distribution of substitution rates with four rate classes, with the α parameter estimated from the data. Shimodaira–Hasegawa tests of alternate topologies (Shimodaira and Hasegawa 1999; Goldman et al. 2000) were performed using Tree-Puzzle (Schmidt et al. 2002).

Gene Characteristics

We calculated the lengths of SGS genes based on annotated gene boundaries. For each mosquito species, we calculated the GC content of the annotated coding regions of each SGS gene and compared these values with the mean GC content of the other protein-coding genes in the genomes. To examine synonymous codon usage in SGS genes compared with other protein-coding genes in mosquito and *Wolbachia* genomes, we used R (<http://www.r-project.org/>) to implement an internal correspondence analysis of codon usage that corrects for differences in amino acid usage between proteins (Lobry and Chessel 2003). We examined the length, location, and conservation of introns annotated in mosquito SGS genes (<http://www.vectorbase.org/>).

Tests for Selection

To test whether *WD0513* is evolving neutrally or under selection, we aligned the nucleotide sequence of this gene with the homologous portion of *WP1346* and *AAEL004181*. We then performed two likelihood ratio tests. First, we compared a model in which all three

branches had the same d_N/d_S ratio with one in which each branch had its own ratio, to determine whether it was appropriate to use branch-specific tests for selection. As the free-ratio model was a significantly better fit to the data ($P > 0.0001$), we then performed a second set of tests, comparing a model in which the d_N/d_S of the *wMel* or *wPip* branch was fixed to 1 with a model in which the d_N/d_S values of all branches were freely estimated. Model likelihoods were estimated using PAML (Yang 1997).

Homologs of Other Genes in the *wMel* Genomic Island

To test whether homologs of any of the other genes in the *wMel* genomic island surrounding *WD0513* were present in *wPip* or in any of the mosquito genomes, we used the amino acid sequences encoded by each of the genes in the island as BLASTP queries against the mosquito gene builds and TBLASTN queries against the genome assemblies of the mosquitoes and *wPip*. As *WD0514* is a gene containing ankyrin repeat domains, which retrieve spurious BLAST hits based on domain sequence similarity rather than overall gene homology, we also used the amino acid sequence of this gene without the ankyrin domains as a query. For the *WD0514* homolog identified in *Ae. aegypti* and the homologs of multiple island genes identified in *wPip*, we confirmed that they were likely orthologs of the query genes by using them as BLAST queries back against the *wMel* protein set and confirming that they were reciprocal best BLAST hits.

Results

We used BLASTP and TBLASTN analyses to identify all SGS homologs present in publicly available sequence databases. There are at least seven SGS genes in the genomes of both *Ae. aegypti* and *Cx. quinquefasciatus* and at least four in *An. gambiae* (supplementary table S1, Supplementary Material online). We also identified a single homologous genomic region in both *wMel* and *wPip* (*WD0512*–*WD0513* and *WP1346*, respectively). No other closely related eukaryotic or prokaryotic homologs were identified in any sequence database; the closest outgroups were relatively distant bacterial proteins (~23% amino acid identity over ~20% of the length of the protein).

Phylogenetic Analysis

We then reconstructed the phylogeny of these genes using Bayesian and maximum likelihood methods (fig. 1; both methods inferred the same topology). If an SGS gene had been horizontally transferred from *Wolbachia* to mosquito, as previously suggested, we would expect the *Wolbachia* genes to cluster at the base of the tree. The earliest branching genes in the ingroup, however, are all from mosquitoes. In contrast, the two *Wolbachia* genes are nested high in the tree within a clade of mosquito genes. Some of the basal ingroup nodes of the tree have only moderate clade confidence and bootstrap values, but this topology is a significantly better fit to the data than one in which the two *Wolbachia* sequences branch at the base of the ingroup ($P = 0.02$, Shimodaira–Hasegawa test).

Because *AAEL004181*, the mosquito gene most closely related to the *Wolbachia* genes, is not nested directly within previously characterized SGS genes (Korochkina et al. 2006) in this tree, we performed a test of genetic distances to be confident that it was a member of this same gene family. We aligned the amino acid sequences of *AAEL004181*, aaSGS1 (encoded by *AAEL009993*), agSGS2 (*AGAP009916*), agSGS4 (*AGAP009917*), and agSGS5 (*AGAP009918*) using T-coffee and then trimmed this alignment with Gblocks as before. We then used Tree-Puzzle to produce a matrix of gamma-corrected maximum likelihood distances between these sequences. Distances between *AAEL004181* and the SGS genes ranged from 1.53 to 1.70 substitutions per residue, whereas distances between SGS genes ranged from 0.51 (aaSGS4–aaSGS5) to 1.75 (aaSGS1–agSGS2). Distances between *AAEL004181* and the SGS genes are smaller than those between some of the previously annotated SGS genes themselves, and we are therefore confident that *AAEL004181* is a member of this gene family.

The two *Wolbachia* genes are not monophyletic in the tree: *WP1346* and *AAEL004181* cluster together, and *WD0513* branches basally to them, possibly indicating that multiple horizontal transfers may have occurred. However, this topology is not a significantly better fit to the data than one in which *AAEL004181* branches outside monophyletic *Wolbachia* sequences ($P = 0.08$, Shimodaira–Hasegawa test), and it is probably most conservative to consider that these three sequences form an unresolved trichotomy.

Gene Characteristics

To further investigate the direction of horizontal transfer, we tested whether *WD0513* and the mosquito SGS genes have characteristics typical of eukaryotic genes (gene length, GC content, and codon usage consistent with the whole mosquito genomes and introns present) or prokaryotic genes (gene length, GC content, and codon usage more similar to that of the *Wolbachia* genome, no introns). Results are summarized in table 1.

The mean length of the mosquito SGS genes is 12,222 bp, which is on the order of the average length of all genes across the three mosquito genomes (8,367 bp). In contrast, *WD0513* is 8,532 bp, over 10 times longer than the mean gene length in the *wMel* genome (852 bp). *WP1346* is 11,646 bp in length, again over 10 times longer than the mean gene length of 944 bp in *wPip*.

Although the GC content of SGS genes is typically lower than the average GC for other genes in the mosquito genomes, they are rarely extreme outliers. Two SGS genes in *Ae. aegypti* have 42% GC content, placing them in the fourth percentile of genes for GC% in the *Ae. aegypti* genome, but many other SGS genes have GC content entirely consistent with the rest of the genome, for example, all of the *Cx. quinquefasciatus* SGS genes (41–50% GC, 21st–53rd percentiles). The GC content of *WD0513*, 35%, is the same as the mean GC across the entire *wMel* genome, and an almost identical pattern is observed for *WP1346* and *wPip*. Similarly, our correspondence analysis also shows that the synonymous codon usage of the SGS genes is

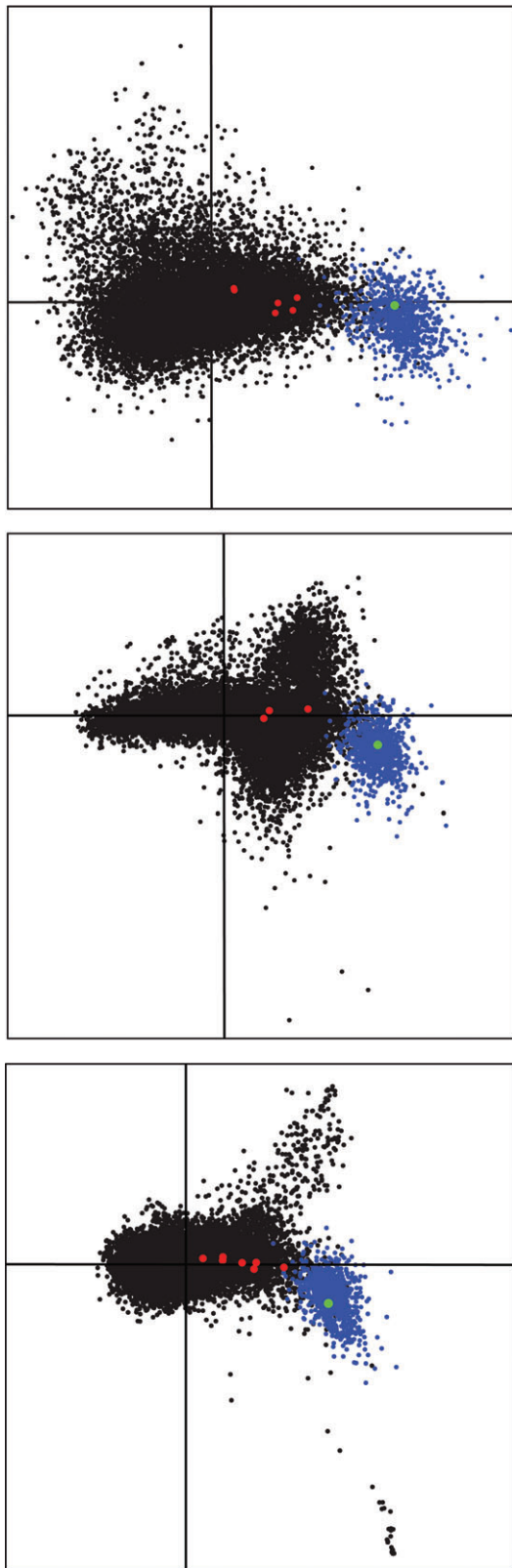


FIG. 2.—First factorial maps for analysis of synonymous codon usage in *wMel* and the three mosquito genomes: (a) *Aedes aegypti*, (b) *Anopheles gambiae*, and (c) *Culex quinquefasciatus*. Each dot on the plots represents a coding sequence (CDS), black for mosquito CDS, red for mosquito SGS proteins, blue for *wMel* CDS, and green for *WD0513*. Red and green dots have been slightly enlarged for better visibility. For

Vectorbase-annotated introns in three sets of closely related SGS gene paralogs, one from each mosquito species, to determine whether we observe either conservation or partial or complete loss of introns in these genes. No such patterns were seen (Supplementary Material online). As these introns are also currently unsupported by expression data, we believe that most or all of the introns annotated in SGS genes are artifactual.

Selection

We investigated whether *WD0513* and *WPI346* are evolving neutrally or under selection by aligning these genes with *AAEL004181* and examining the ratio of non-synonymous to synonymous changes (d_N/d_S) along each branch of the triplet. *WD0513* has a d_N/d_S of 0.17, and *WPI346* has a d_N/d_S of 0.14. Both these values are significantly lower than a neutral d_N/d_S of 1 ($P > 0.0001$, likelihood ratio test, Bonferroni correction for multiple tests), indicating that these genes are probably evolving under purifying selection. Recombination (as is known to occur between strains of *Wolbachia*, Baldo et al. 2005) can affect analyses of selection (Anisimova et al. 2003). However, recombination tends to bias these analyses toward detection of positive rather than purifying selection, and so our estimate of d_N/d_S is likely to be conservative.

The genomic region in *wPip* homologous to the 5' end of *WD0513* has been disrupted by the insertion of a transposable element of the IS110 family, and the start codon of *WPI346* corresponds to codon 154 in *WD0513*. It is possible that this slightly truncated protein remains functional in *wPip*, and the low d_N/d_S value for this gene reflects current purifying selection. Alternatively, the IS110 element may be a recent insertion that has only recently disrupted this gene; such recent activity would be consistent with the fact that the three copies of this element in the *wPip* genome are identical to one another at the nucleotide level. Nonetheless, the absence of stop codons in *WD0513* and *WPI346*, together with the low estimated values of d_N/d_S , offers further evidence for the action of long-term purifying selection on these genes.

Multiple Genes Transferred from Mosquito to *Wolbachia*

WD0513 in *wMel* is part of a larger genomic island of 13 genes (*WD0506*–*WD0518*) that is present in *wMel* and two closely related A-group strains of *Wolbachia* but absent from other A- and B-group strains examined (Iturbe-Ormaetxe et al. 2005). Given this unusual taxonomic distribution, and the presence of multiple transposable element genes in the island, we tested whether other island genes might also have originated by HGT from mosquito. Most genes in the genomic island had either no mosquito homologs or had many closer prokaryotic than eukaryotic

←

each mosquito species, the codon usage of the SGS genes is well within the range expected for the mosquito genome, and the codon usage of *WD0513* is also consistent with that of the *wMel* genome.

homologs, indicating that no interdomain HGT of these genes has taken place.

We did, however, identify likely ancestors of *WD0512* and *WD0514* in the mosquito genomes. The 3' end or the 3'-UTR of every mosquito SGS gene we identified is homologous to the first 300–500 bp of *WD0512*, and the 3'-UTR of *AAEL004181* is homologous to the full length of *WD0512*. *WD0514* has a single mosquito homolog, *AAEL004188*, which is located directly upstream of *AAEL004181* in the *Ae. aegypti* genome. The presence of full-length homologs of *WD0512* and *WD0514* on either side of *AAEL004181* offers further support for the close phylogenetic relationship between this SGS gene and *WD0513* (fig. 1).

Horizontal Transfer between *Wolbachia* Strains

Orthologs of numerous genes from the genomic island were found clustered in the *wPip* genome. It seems likely that the entire island was at one time present in the *wPip* genome but that it has subsequently been subject to partial or complete duplication followed by gene loss (data not shown). This island is absent from all A- and B-group *Wolbachia* tested other than the closely related strains *wMel*, *wMelCS*, and *wMelPop* (A-group) (Iturbe-Ormaetxe et al. 2005) and *wPip* (B-group). This discontinuous taxonomic distribution suggests either that the island was present in the ancestor of A-group and B-group *Wolbachia* and has subsequently been lost in the majority of sampled *Wolbachia* strains or that it has been horizontally transferred between strains more recently. Several lines of evidence support interstrain HGT as a more likely explanation. First, the *wMel* and *wPip* orthologs of island genes have anomalously high sequence similarity (up to 96% nucleotide identity for each ortholog pair's longest BLASTN high-scoring segment pair) compared with the rest of the two genomes. Furthermore, numerous independent losses would be required to produce the genomic island's current taxonomic distribution if it had been present in the ancestor of A- and B-group *Wolbachia* (Iturbe-Ormaetxe et al. 2005). Finally, mechanisms for genetic transfer between A- and B-group *Wolbachia* clearly exist, given the previously reported examples of intergroup HGT (Masui et al. 2000; Bordenstein and Wernegreen 2004).

Discussion

Our analyses support the claim of Korochkina et al. (2006) that the sequence similarity observed between the SGS genes and *WD0513* is due to an HGT event. Genes may have similar sequences due either to common origin or to convergence, and there are two reasons to believe that the data examined here are not an example of convergence. First, convergent similarity is likely to be restricted to relatively short functional regions of genes rather than complete gene sequences (Eisen 1998). In contrast to this, we observe sequence similarity along the full lengths of *WD0513* and *AAEL004181*. Secondly, our putatively transferred sequence fragment contains multiple genes: *WD0512* and *WD0514*, the genes flanking *WD0513* in the *wMel* genome, show strong similarity to the 3'-UTR of *AAEL004181* and to the upstream *Ae. aegypti* gene *AAEL004188*, respec-

tively. Convergence across the lengths of multiple neighboring genes and noncoding regions seems extremely unlikely.

The direction in which this horizontal transfer occurred is less clear, but we believe that the balance of evidence (table 1) supports a transfer from a eukaryotic host (probably mosquito) to *Wolbachia*. SGS gene homologs have been identified in every mosquito species for which we have whole genomes or salivary gland expression libraries. In contrast, homologs have been found in only a minority of the *Wolbachia* strains sampled (Iturbe-Ormaetxe et al. 2005). This sparse distribution in *Wolbachia* suggests that it was the recipient rather than the donor of the transferred gene (Andersson 2005). This direction of transfer is also strongly implied by the phylogeny of these genes (fig. 1), as the basal position of mosquito sequences, and the nesting of the *Wolbachia* homologs within a mosquito clade, is supported both by multiple tree reconstruction methods and by a maximum likelihood test of alternative topologies.

Finally, a direction of transfer from eukaryote to prokaryote is also supported by the lengths of the SGS genes and their *Wolbachia* homologs. The mosquito SGS genes are relatively long even for the eukaryotic genomes in which they are found; even the shortest of them would be exceptionally long for a gene derived from a bacterium. *WD0513*, at 8,532 bp, is the longest open reading frame in the *wMel* genome and an extreme outlier in terms of length. Only one other *wMel* gene, *WD0024*, which is a fusion of two RNA polymerase subunit genes, approaches this length; the third longest gene, by comparison, is only 4,725 bp. The mean gene length in *wMel* is only 852 bp, less than one-tenth of the length of *WD0513*. Similarly, *WP1346* is 11,649 bp long, over 10 times as long as the mean gene length of 944 bp in *wPip*. In this regard, *WD0513* and *WP1346* appear to be more like typical eukaryotic than prokaryotic genes.

If *WD0513* has been transferred from a eukaryotic host, we might expect it to retain other characteristics of its native genome, such as GC content and patterns of synonymous codon usage. *WD0513*, however, does not differ from the rest of the *wMel* genome for these traits. The SGS genes do not show any consistent deviation from the GC content or codon usage of the mosquito genomes either, and these traits are therefore uninformative for inferring the direction of transfer. This lack of signal is presumably due to amelioration of the original base composition and codon usage of the transferred sequences to reflect the mutational biases and/or selective pressures of the recipient genome. As this transfer event was an ancient one (*WD0513* is only ~50% identical at the amino acid level to *AAEL004181*), there has been substantial evolutionary time for this amelioration to take place.

Only one set of data potentially supports a transfer in the opposite direction, from *Wolbachia* to host. It is rare for mosquito genes to lack introns—90% of *Ae. aegypti* genes have introns, and the proportion rises to almost 94% in *An. gambiae* (Nene et al. 2007). If our assessment is correct, and the introns currently annotated in SGS genes are artificial, then horizontal transfer of these genes from a bacterial source is a possible explanation for their lack of introns. On the other hand, very few intronless eukaryotic genes have arisen from HGT: most are thought to have been generated by retrotransposition followed by genomic

reintegration or other mechanisms requiring only vertical descent (Sakharkar et al. 2006). This indication of directionality is therefore not compelling.

Recent findings of substantial HGT between prokaryotic endosymbionts and their eukaryotic hosts (Kondo et al. 2002; Dunning Hotopp et al. 2007; Nikoh et al. 2008) indicate that interdomain gene transfers may be unexpectedly frequent. Such transfers have been assumed to be unidirectional, from prokaryote to eukaryote, but multiple lines of evidence presented here support the conclusion that *WD0513* is the result of eukaryote-to-prokaryote HGT. Gene transfers in this direction are far less common than from prokaryote to eukaryote (Andersson 2005), but given the frequency with which hosts and endosymbionts exchange genetic material over evolutionary time scales (Dunning Hotopp et al. 2007), it is entirely possible that one or more host-to-*Wolbachia* transfers may have occurred. Host genomes may therefore be occasional, but important, sources of novel genetic material for their bacterial endosymbionts.

Regardless of the direction in which this HGT occurred, it remains an event of strong evolutionary interest. Unlike many of the interdomain transfers previously described (Dunning Hotopp et al. 2007; Nikoh et al. 2008), the genetic distance between *WD0513* and the SGS genes indicates that the transferred copy has been retained over a substantial period of evolutionary time, and both donor and recipient genes appear to be functional. Some members of the SGS gene family have been previously characterized (Korochkina et al. 2006), and although the function of *WD0513* is currently unknown, it has diverged substantially from its mosquito homologs without pseudogenization, is evolving under purifying selection, and is expressed (Iturbe-Ormaetxe et al. 2005), indicating that it probably plays some role in the biology of the *Wolbachia* strains in which it is found. This HGT thus fulfills the criteria of evolutionary significance (Blaxter 2007), demonstrating that interdomain HGTs are potentially powerful sources of evolutionary innovation.

Supplementary Material

Supplementary table S1 and supplementary figure S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Marie Sémon, Lyn Cook, Mike Crisp, Mark Ragan, Steve Sinkins, members of the O'Neill and McGraw labs, and attendees of the fifth International *Wolbachia* Conference for valuable discussion and comments on the analyses. This work was supported by grants from the Australian Research Council (ARC; DP0557987) and National Health and Medical Research Council (NHMRC; 496601), and a UQ postdoctoral fellowship to M.W.

Literature Cited

Andersson JO. 2005. Lateral gene transfer in eukaryotes. *CMLS—Cell Mol Life Sci.* 62:1182–1197.

- Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics.* 164:1229–1236.
- Arca B, Lombardo F, Francischetti IMB, Pham VM, Mestres-Simon M, Andersen JF, Ribeiro JMC. 2007. An insight into the sialome of the adult female mosquito *Aedes albopictus*. *Insect Biochem Mol Biol.* 37:107–127.
- Arca B, Lombardo F, Valenzuela JG, Francischetti IMB, Marinotti O, Coluzzi M, Ribeiro JAC. 2005. An updated catalogue of salivary gland transcripts in the adult female mosquito, *Anopheles gambiae*. *J Exp Biol.* 208:3971–3986.
- Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. 2005. Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol.* 23:437–449.
- Blaxter M. 2007. Symbiont genes in host genomes: fragments with a future? *Cell Host Microbe.* 2:211–213.
- Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol.* 21:1981–1991.
- Brennan JDG, Kent M, Dhar R, Fujioka H, Kumar N. 2000. *Anopheles gambiae* salivary gland proteins as putative targets for blocking transmission of malaria parasites. *Proc Natl Acad Sci USA.* 97:13859–13864.
- Calvo E, Dao A, Pham VM, Ribeiro JMC. 2007. An insight into the sialome of *Anopheles funestus* reveals an emerging pattern in anopheline salivary protein families. *Insect Biochem Mol Biol.* 37:164–175.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17:540–552.
- Dunning Hotopp JC, Clark ME, Oliveira DCSG, et al. (20 co-authors). 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science.* 317:1753–1756.
- Eisen JA. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8:163–167.
- Fenn K, Conlon C, Jones M, Quail MA, Holroyd NE, Parkhill J, Blaxter M. 2006. Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. *PLoS Pathog.* 2:887–899.
- Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst Biol.* 49:652–670.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. 2008. How many species are infected with *Wolbachia*?—a statistical analysis of current data. *FEMS Microbiol Lett.* 281:215–220.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.
- Iturbe-Ormaetxe I, Burke GR, Rieger M, O'Neil SL. 2005. Distribution, expression and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol.* 187:5136–5145.
- Klasson L, Walker T, Sebahia M, et al. (12 co-authors). 2008. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol Biol Evol.* 25:1877–1887.
- Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T. 2002. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc Natl Acad Sci USA.* 99:14280–14285.
- Korochkina S, Barreau C, Pradel G, Jeffery E, Li J, Natarajan R, Shabanowitz J, Hunt D, Frevert U, Vernick KD. 2006. A mosquito-specific protein family includes candidate receptors for malaria sporozoite invasion of salivary glands. *Cell Microbiol.* 8:163–175.

- Lobry JR, Chessel D. 2003. Internal correspondence analysis of codon and amino-acid usage in thermophilic bacteria. *J Appl Genet.* 44:235–261.
- Masui S, Kamoda S, Sasaki T, Ishikawa H. 2000. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J Mol Evol.* 51:491–497.
- Nene V, Wortman JR, Lawson D, et al. (95 co-authors). 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science.* 316:1718–1723.
- Nikoh N, Tanaka K, Shibata F, Kondo N, Hizume M, Shimada M, Fukatsu T. 2008. *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res.* 18:272–280.
- Notredame C, Higgins DG, Heringa J. 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302:205–217.
- Rambaut A. 1996. Se-AL: sequence alignment editor v2.0. Available from: <http://evolve.zoo.ox.ac.uk/>.
- Sakharkar KR, Sakharkar MK, Culiati CT, Chow VTK, Pervaiz S. 2006. Functional and evolutionary analyses on expressed intronless genes in the mouse genome. *FEBS Lett.* 580:1472–1478.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics.* 18:502–504.
- Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 16:1114–1116.
- Werren JH, Windsor DM. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc R Soc Lond Ser B Biol Sci.* 267:1277–1285.
- Wu M, Sun LV, Vamathevan J, et al. (30 co-authors). 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* 2:327–341.
- Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13: 555–556.

Laura Katz, Associate Editor

Accepted November 1, 2008