Insight into the Transmission Biology and Species-Specific Functional Capabilities of Tsetse (Diptera: Glossinidae) Obligate Symbiont *Wigglesworthia*

Rita V. M. Rio,[a] Rebecca E. Symula,[b] Jingwen Wang,[c] Claudia Lohs,[c*] Yi-neng Wu,[c] Anna K. Snyder,[a] Robert D. Bjornson,[d] Kenshiro Oshima,[e] Bryan S. Biehl,[f] Nicole T. Perna,[f] Masahira Hattori,[e] and Serap Aksoy[c]

Department of Biology, West Virginia University, Morgantown, West Virginia, USA[a]; Department of Ecology and Evolutionary Biology, Yale University, New Haven Connecticut, USA[b]; Division of Epidemiology of Microbial Diseases, School of Public Health, Yale University, New Haven, Connecticut, USA[c]; Department of Computer Science and W. M. Keck Biotechnology Resource Laboratory, Yale University, New Haven, Connecticut, USA[d]; Graduate School of Frontier Sciences, The University of Tokyo, Kashiwanoha, Kashiwa, Chiba, Japan[e]; and Department of Genetics, University of Wisconsin—Madison, Madison, Wisconsin, USA[f]

ABSTRACT Ancient endosymbionts have been associated with extreme genome structural stability with little differentiation in gene inventory between sister species. Tsetse flies (Diptera: Glossinidae) harbor an obligate endosymbiont, *Wigglesworthia*, which has coevolved with the *Glossina* radiation. We report on the ~720-kb *Wigglesworthia* genome and its associated plasmid from *Glossina morsitans morsitans* and compare them to those of the symbiont from *Glossina brevipalpis*. While there was overall high synteny between the two genomes, a large inversion was noted. Furthermore, symbiont transcriptional analyses demonstrated host tissue and development-specific gene expression supporting robust transcriptional regulation in *Wigglesworthia*, an unprecedented observation in other obligate mutualist endosymbionts. Expression and immunohistochemistry confirmed the role of flagella during the vertical transmission process from mother to intrauterine progeny. The expression of nutrient provisioning genes (*thiC* and *hemH*) suggests that *Wigglesworthia* may function in dietary supplementation tailored toward host development. Furthermore, despite extensive conservation, unique genes were identified within both symbiont genomes that may result in distinct metabolomes impacting host physiology. One of these differences involves the chorismate, phenylalanine, and folate biosynthetic pathways, which are uniquely present in *Wigglesworthia morsitans*. Interestingly, African trypanosomes are auxotrophs for phenylalanine and folate and salvage both exogenously. It is possible that *W. morsitans* contributes to the higher parasite susceptibility of its host species.

IMPORTANCE Genomic stasis has historically been associated with obligate endosymbionts and their sister species. Here we characterize the *Wigglesworthia* genome of the tsetse fly species *Glossina morsitans* and compare it to its sister genome within *G. brevipalpis*. The similarity and variation between the genomes enabled specific hypotheses regarding functional biology. Expression analyses indicate significant levels of transcriptional regulation and support development- and tissue-specific functional roles for the symbiosis previously not observed in obligate mutualist symbionts. Retention of the genetically expensive flagella within these small genomes was demonstrated to be significant in symbiont transmission and tailored to the unique tsetse fly reproductive biology. Distinctions in metabolomes were also observed. We speculate an additional role for *Wigglesworthia* symbiosis where infections with pathogenic trypanosomes may depend upon symbiont species-specific metabolic products and thus influence the vector competence traits of different tsetse fly host species.

Address correspondence to Serap Aksoy, serap.aksoy@yale.edu.

*Present address: Max F. Perutz Laboratories, University of Vienna, Vienna, Austria.

R. V. M. Rio and R. E. Symula contributed equally to this article.

Microbial symbioses with eukaryotic hosts are ubiquitous. Despite the wide prevalence of symbiosis, obligate associations, where partners are inextricable from and entirely dependent on one another, are relatively rare. Endosymbionts residing within host cells are typically vertically transmitted between generations with high fidelity, coupling partners, and this results in an intimate, specialized association over evolutionary time. Such symbiotic associations can enable hosts to acquire new metabolic capabilities and thus thrive in novel niches. This idea was first hypothesized for insects that subsist on nutrient-poor phloem sap, where symbionts supplement dietary deficiencies (1). Similar nutritional symbioses have since been identified in many insects with nutritionally restricted diets, including tsetse flies, carpenter ants, and many plant-feeding insects (2). Based on whole-genome sequences, the obligate symbiont genomes have all been drastically reduced in size in comparison to those of their free-living
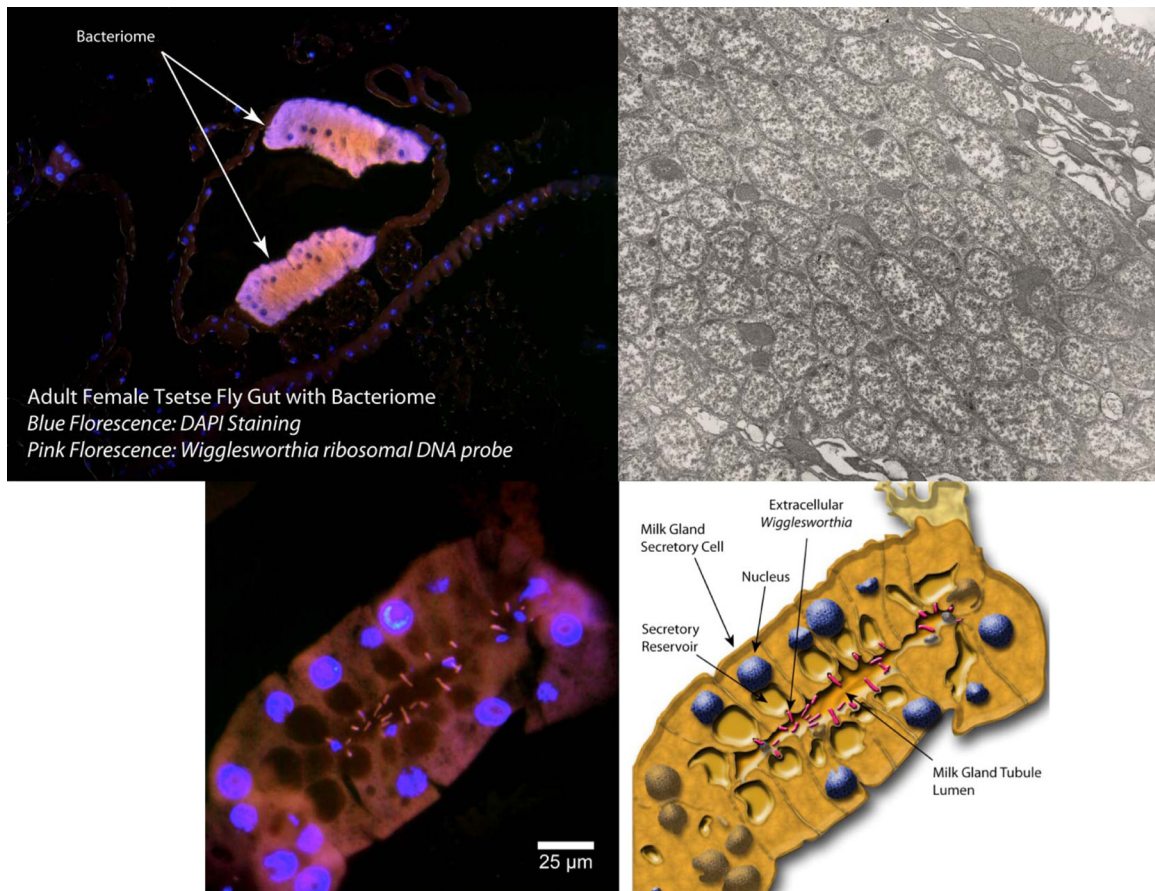
FIG 1 Localization of *Wigglesworthia* within the tsetse fly. *Wigglesworthia* resides within the bacteriome organ (top left) and is found free in the cytoplasm of specialized cells known as bacteriocytes (top right). (Bottom) *Wigglesworthia* is present extracellularly in the milk gland tissue. (Bottom left) Fluorescence *in situ* hybridization (FISH) staining for *Wigglesworthia*. (Bottom right) Schematic drawing based on FISH results shown on the left. DAPI staining indicates nuclei in bacteriocytes and milk gland tubule cells. Pink fluorescent rhodamine staining shows *Wigglesworthia* within bacteriocytes and in milk gland lumen.

relatives and display high A+T bias (3, 4). These traits are thought to have arisen through relaxed natural selection and resulting genome deterioration (5). The obligate symbionts of aphids, carpenter ants, and tsetse flies, *Buchnera*, *Blochmannia*, and *Wigglesworthia*, respectively, form a close lineage in the gamma-proteobacteria and are believed to have independently established host relations. Despite their close relatedness, the extant genomes of the symbionts have undergone drastic, yet distinct, adaptive reductions. It is thought that genes retained in each small genome are necessary for functional capabilities that complement host physiology and ecology, with gene inventory having some specificity to the host lineage (5).

Every tsetse fly species is associated with a distinct *Wigglesworthia glossinidia* lineage (6). Phylogenetic analysis of *Wigglesworthia* shows concordant history between symbiont lineages and their host species, indicating partner coevolution. Further, molecular clock methods suggest that this symbiosis is about 80 million years old (6). In addition to *Wigglesworthia*, tsetse fly laboratory colonies and some natural populations can harbor a commensal bacterial symbiont, *Sodalis glossinidius*, of relatively recent establishment (7, 8) and parasitic *Wolbachia* infections (9, 10).

Tsetse flies feed exclusively on nutrient-poor vertebrate blood. Unlike many other insects, tsetse flies display viviparous repro-duction (deposition of live late-stage larvae rather than eggs), where the mother develops a single oocyte at a time and then carries and nourishes the resulting embryo and larva in an intra-uterine environment. The mother undergoes parturition to a fully developed larva that quickly pupates and remains dormant for about a month prior to adult metamorphosis. Thus, throughout its developmental cycle, the tsetse fly is solely dependent on its vertebrate host blood diet. The obligate mutualist *Wigglesworthia* is thought to complement the exclusive blood diet of its host.

Wigglesworthia resides intracellularly in bacteriocytes, which form the bacteriome organ in the anterior midgut (see Fig. 1). In the bacteriocyte cytoplasm, the symbionts live free and are not surrounded by host membranes. In addition to the bacteriome, extracellular *Wigglesworthia* is also detected in the milk gland lu-men (11, 12). Extracellular *Wigglesworthia*, along with *Sodalis*, is maternally transmitted to the tsetse fly's intrauterine progeny through milk secretions synthesized via the modified accessory gland (milk gland) that connects to the uterus (13). Without *Wigglesworthia*, tsetse fly females are reproductively sterile. Given that the exclusive tsetse fly blood diet is low in vitamins, coupled with data from dietary supplementation experiments of antibiotic-fed (symbiont-free) tsetse flies, a putative role in vita-min metabolism has been suggested for the symbionts (14). In

addition to host dietary supplementation through vitamin provisioning, the presence of *Wigglesworthia* is essential for the tsetse fly's immune system maturation. It has been possible to develop flies that lack *Wigglesworthia* by maintaining fertile tsetse fly females on ampicillin-supplemented blood meals (11). This is because the antibiotic ampicillin does not affect the intracellular forms of *Wigglesworthia* within the bacteriome but can clear the extracellular *Wigglesworthia* from tsetse fly milk. The resulting progeny of ampicillin-treated females lack *Wigglesworthia* (Gmm^{Wig-}) but retain the commensal *Sodalis*. In comparison to their normal counterparts, Gmm^{Wig-} adults are highly susceptible to trypanosome midgut infections (11) and microbial challenge (15). Thus, it appears that when larvae develop without *Wigglesworthia*, cellular immunity is particularly compromised in the emerging adult progeny (15).

The genome of *Wigglesworthia glossinidia* characterized from *Glossina brevipalpis* (referred to here as WGB) is about 697 kb in size and has a small plasmid (pWgb). The genome encodes 621 predicted coding sequences (CDSs) and displays a high (82%) adenine-thymine (A+T) bias (16, 17). It is possible that the high A+T content of *Wigglesworthia* resulted from the loss of repair and recombination functions such as the SOS, base excision, and nucleotide excision repair system (*uvrABC*). Surprisingly, the important gene coding for the DNA replication initiation protein DnaA was missing from the WGB genome—an observation previously unprecedented in eubacteria. More than 10% of the retained CDSs are involved in the biosynthesis of cofactors, prosthetic groups, and carriers, supporting *Wigglesworthia*'s genetic contributions in the *de novo* metabolism of biotin, thiazole, lipoic acid, flavin adenine dinucleotide (riboflavin, B$_2$), folate, pantothenate, thiamine (B$_1$), pyridoxine (B$_6$), and protoheme and further substantiating the role of *Wigglesworthia* in host dietary supplementation (14). In addition to providing its tsetse fly host with vitamins, comparative genome studies with *Sodalis* indicate that *Wigglesworthia* may also provide thiamine to *Sodalis*, which lacks the thiamine biosynthetic pathway but has retained the transporter for acquiring thiamine (18). The functional complementation of symbiont genomes has been postulated to reduce competition between microbes, as well as prevent the possibility of symbiont replacement especially during the early establishment of a dual symbiosis (2).

Here we describe the genome of a second *Wigglesworthia* species isolated from *Glossina morsitans morsitans* (referred to here as WGM). Phylogenetic molecular clock analyses suggest that tsetse fly host species of WGB and WGM have been distinct for 50 to 80 million years (6). We compare the genome structures and gene inventories of WGM and WGB and explore evolutionary patterns in the genes which may contribute to functional variation within their respective tsetse fly host species. We describe the expression of *Wigglesworthia* genes that may be significant for tsetse fly nutrition through development. Lastly, we provide support for the role of flagella during the crucial symbiont maternal transmission process. We discuss similarities and differences between the two genomes that may ultimately affect important host physiological processes, including varying vector competence of the tsetse fly host species.

RESULTS

Features of the WGM genome. The genome of WGM consists of a circular chromosome of 719,535 bp (with a guanine-plus-

cytosine [G+C] content of 25%) and a single plasmid of 5,198 bp. The putative origin of replication, without a clear G+C skewing and diagnostic DnaA boxes, was assigned to the same A+T-rich region upstream of the *gidA* locus as WGB (Fig. 2A and B). Table 1 summarizes the general features of the WGM genome, relative to those of other insect endosymbionts, including that of WGB, the relatively recently established genus *Sodalis*, the related ancient obligate symbiont, *Blochmannia*, of carpenter ants, and two *Buchnera* symbionts from different aphid hosts, respectively (7, 16, 19–21). Both the genome size and the exceptionally low G+C content of WGM were comparable to those reported for other ancient endosymbionts, including WGB. Annotation revealed that, similar to the other small obligate genomes, the coding content of WGM is high (83.9%) with 620 predicted CDSs at an average length of 979 bp (Table 1). The high A+T bias of the WGM chromosome was reflected in the higher average predicted isoelectric points of putative proteins, as was noted in WGB (9.84 in obligates versus 7.2 in *Sodalis*). Like WGB, WGM has two identical copies of each of the rRNA genes (*rrsH* and *rrlB*) (Fig. 2B). Similar to those of other symbionts, these rRNA genes have higher G+C contents, 49.3% and 45.8%, respectively, than protein coding genes. A total of 11 recognizable pseudogenes were identified in WGM (see Table S1 in the supplemental material), and these were distributed throughout the genome.

Comparison of WGM and WGB genome structures. Alignment of the two *Wigglesworthia* genomes indicates high chromosomal synteny, as was previously described for the *Buchnera* (22) and *Blochmannia* (23) genomes. However, since the divergence of WGB and WGM, a chromosomal inversion has occurred in one of the lineages (Fig. 2A). The inversion, which can be interpreted as an either 550-kb or 170-kb inversion due to the circular chromosome of *Wigglesworthia*, occurs approximately 150 kb from the *gidA* locus, in proximity to the origin of replication (Fig. 2B). An inversion in proximity to the origin of replication could create imbalanced replichores between the *Wigglesworthia* genomes. The inversion is flanked on either end by the rRNA genes *rrsH* and *rrlB*. In both *Wigglesworthia* genomes, within this G+C-rich region, and specifically within the *rrsH* gene, is a sequence that is nearly identical to the *Escherichia coli* Chi recombination hot spot (24). The sequence differs by only a single base (in bold): *E. coli*, **G**CTGGTGG; *Wigglesworthia*, **T**CTGGTGG. Since the RecA protein, which has been retained in both *Wigglesworthia* genomes, has the highest affinity to the TGG repeats in *E. coli* (25), we propose that this site (~480 bp into *rrsH*) likely demarcates the inversion site.

Comparative analysis of WGB and WGM contents. Comparative analyses of the WGB and WGM genomes reveal a shared set of 599 CDSs (Fig. 3A). Both genomes have retained pathways involved in B vitamin biosynthesis, including biotin (B$_7$), thiazole (B$_1$), riboflavin (B$_2$), pantothenate (B$_5$), and pyridoxine (B$_6$). However, genetic components involved in the synthesis of cobalamin (B$_{12}$) and nicotinate (B$_3$) appear to be absent from WGM and WGB. Genes necessary for the synthesis of a complete flagellum apparatus have also been preserved in both genomes. Since the release of the WGB genome annotation, genes exhibiting high sequence identity to the Wg001 to Wg003 orphan genes have been reported in other host-associated bacteria. Wg001 to Wg003 are homologous to a putative transmembrane protein, an endonu-
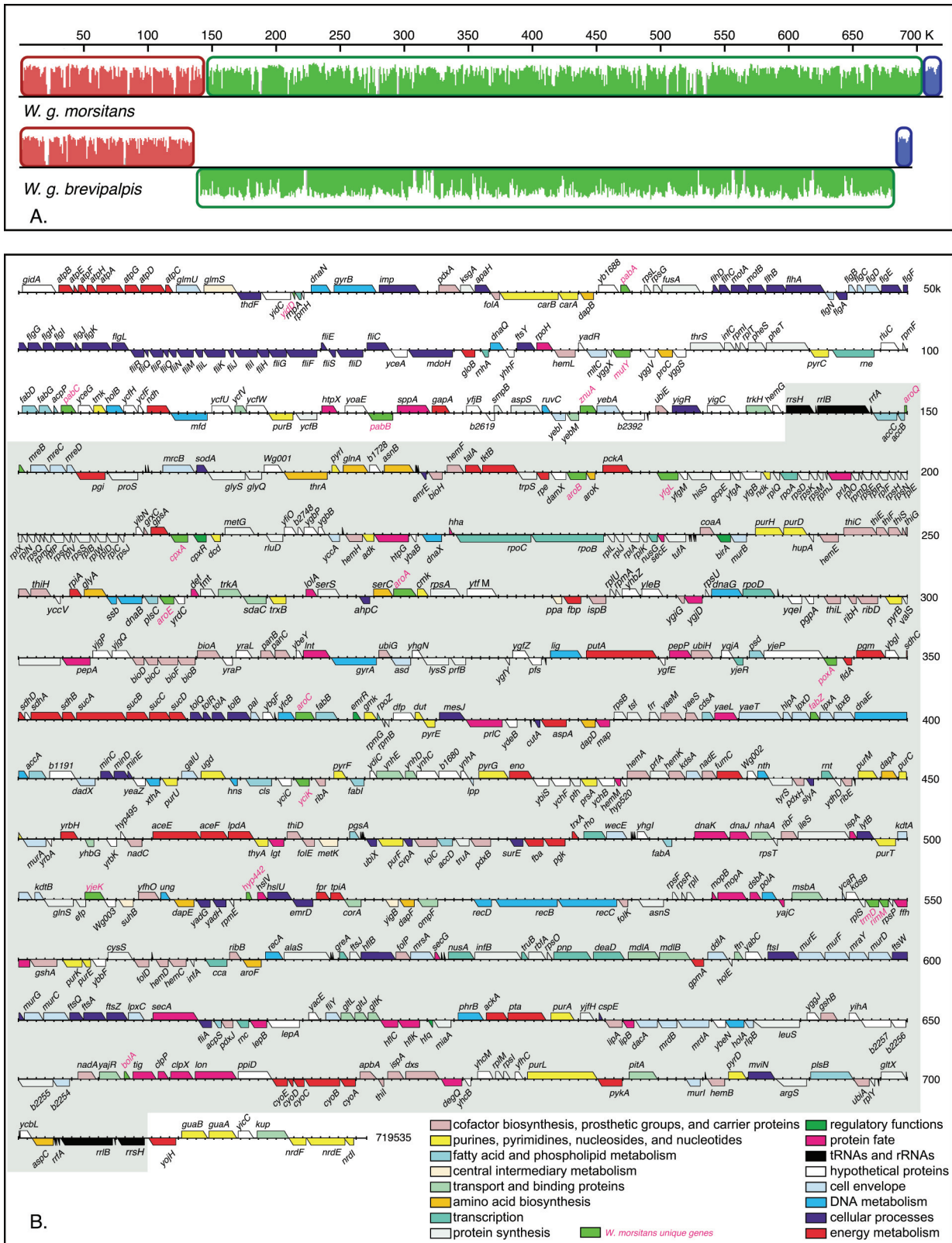
FIG 2 Linearized comparison of the genomes of WGB and WGM. (A) Genome alignment as represented from Mauve. Each colored block represents a locally collinear block (LCB) of DNA that has not undergone rearrangement within its boundaries. Bar height indicates average nucleotide similarity within a region. The green LCB is inverted, as indicated by the relative reverse orientation of the block in each genome. (B) Annotation of the WGM genome. Each gene is shown as a trapezoid with the straight side representing the start codon. Genes above the line are on the positive strand, while genes below the line are on the negative strand. The shaded area indicates the inverted region relative to the WGB genome. Genes are labeled and color coded according to the different functional categories assigned.

TABLE 1 Comparison of genome features of various insect bacterial endosymbionts

Species described	Chromosome size (bp)	No. of plasmids [size (bp)]	Protein coding (%)	G+C content (%)	No. of predicted CDSs	No. of rRNA operons	tRNA	No. of pseudogenes	Average ORF length (bp)
Wigglesworthia spp.									
W. morsitans	719,535	1 (5,198)	83.9	25	620	2	34	11	979
W. brevipalpis[a]	697,742	1 (5,280)	89	22.5	618	2	34	8	988
Sodalis glossinidius[b]	4,171,146	3 (121,356)	50.9	54.7	2,432	7	69	972	873
Blochmannia floridanus[c]	705,557	0	83.2	27.4	583	1	37	6	1,007
Buchnera strains									
BAp[e]	640,681	1 (≥15,044)	88	26.3	583	1	32	12	988
BSg[f]	641,454	2 (~22,267)[g]	84.5	26.2	545	1	38	33	978

[a] As reported by Akman et al. (16) and with our own revision.
[b] As reported by Toh et al. (7).
[c] As reported by Gil et al. (21).
[d] *Buchnera aphidicola* strains BAp and BSg were selected because of similar 16S rRNA genetic distances between isolates relative to *W. morsitans* and *W. brevipalpis*.
[e] *B. aphidicola* strain isolated from *Acyrthosiphon pisum* as reported by Shigenobu et al. (19).
[f] *B. aphidicola* strain isolated from *Schizaphis graminum* as reported by Tamas et al. (20).
[g] Tryptophan plasmid size as reported by Lai et al. (62).

cleave, and an integral membrane protein, respectively. Notably, these genes have also been retained in WGM. Unlike most bacteria, WGM and WGB both lack *dnaA*, suggesting gene loss in the ancestral lineage prior to host diversification.

Unique genes (i.e., those lacking in the sister genome) were identified in WGM and WGB (Fig. 3B compares the unique gene inventories). Notably, significant differences in the distribution of functional categories of unique genes were observed between WGB and WGM (Kolmogorov-Smirnov test, $\alpha = 0.05$). These 19 and 21 genes, respectively, and their putative biological roles are listed in Table S2 in the supplemental material. In addition, the positions of these genes within the WGM genome are highlighted in Fig. 2. The retention of these unique

genes suggests metabolic distinctions within the proteomes of the two *Wigglesworthia* sister species. Notably, analysis of the WGM-specific gene set reveals the presence of a complete shikimate biosynthetic pathway in which 3-deoxy-d-arabino-heptulosonate-7-phosphate can be converted into chorismate (see Fig. S2 in the supplemental material). This pathway is degraded in the WGB genome, where only *aroG* and *aroK* homologs are still detectable. Acting downstream of the chorismate pathway, WGM also contains *pabA*, *pabB* (encoding aminodeoxychorismate synthases II and I, respectively), and *pabC* (encoding 4-amino-4-deoxychorismate lyase), which catalyzes the reaction from chorismate to *p*-aminobenzoate, an essential component in folate biosynthesis (see Fig. S2 in the supplemental material). The WGM genome also contains an *aspC* homolog that can also be used following chorismate biosynthesis toward phenylalanine production.
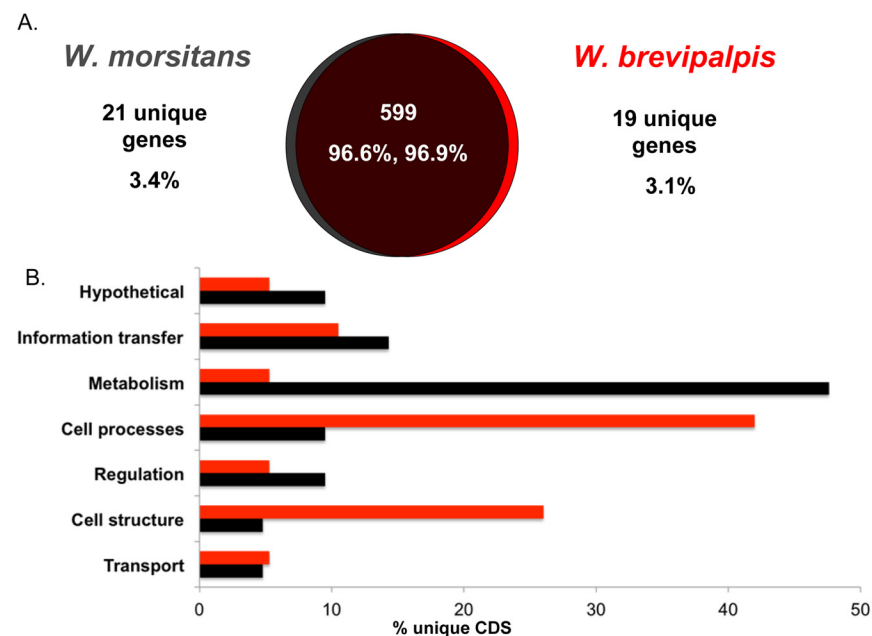
WGM plasmid and putative functions. The WGM plasmid (pWgm) is 5,198 bp in length with a G+C content of 24%. pWgm has six CDSs (see Fig. S1 in the supplemental material), which are homologous to the six CDSs previously identified in pWgb (16). The four CDSs, with only a minor frequency of indels, encode a spermidine acetyltransferase (pWgm open reading frame 1 [ORF1], 177 amino acids [aa]; WGpWb0002, 174 aa), a putative mechanosensitive channel protein (encoded by *yggB*; pWgm ORF3, 282 aa; WGpWb0005, 280 aa), a putative heat shock protein (pWgm ORF5, 137 aa; WGpWb0006, 133 aa), and a conjugative transfer surface exclusion lipoprotein (pWgm ORF6, 243 aa; WGpWb0003, 239 aa). The remaining two CDSs, which exhibit higher sequence variation, encode a replication protein A,

FIG 3 Comparative analyses of the WGM and WGB genomes. (A) The *Wigglesworthia* pangenome consists of a core 599 CDSs with an additional 21 and 19 unique CDSs within the WGM and WGB genomes, respectively. (B) Distribution of unique CDSs present per functional category.
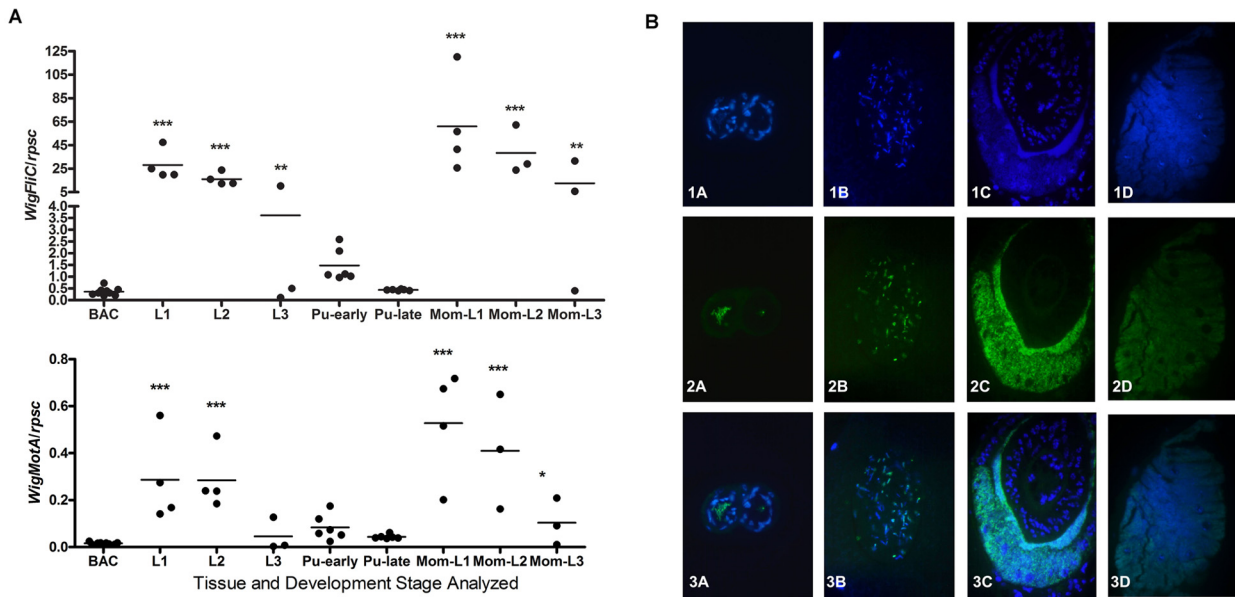
FIG 4 *Wigglesworthia* flagella are utilized during maternal transmission. (A) Normalized qRT-PCR-based gene expression results for the *fliC* and *motA* genes in the bacteriome (BAC), different stages of intrauterine larvae (L1, L2, and L3), newly deposited pupae (Pu-early), late pupae (Pu-late), and carcasses of mothers that carry the corresponding intrauterine larvae (Mom-L1, Mom-L2, and Mom-L3). Asterisks indicate statistically significant differences between the bacteriome and various developmental stages. ***, $P < 0.0001$; **, $P < 0.001$; *, $P < 0.05$. (B) Images of WGM FliC-specific antibody staining on cross sections of the tsetse fly common milk duct (column A), duct within the larval gut (column B), larval bacteriome (column C), and adult bacteriome (column D). Row 1 represents DAPI staining, row 2 represents FliC antibody staining, and row 3 represents the merged images of rows 1 and 2.

RepA (pWgm ORF2, 239 aa), that has a 36-aa deletion at the 5′ end in comparison to its pWgb homolog, and a hypothetical protein (pWgm ORF4, 311 aa) that contains 12 nonsynonymous changes occurring within the first 20 aa of the sequence relative to its pWgb homolog.

Maternal transmission of *Wigglesworthia* to intrauterine progeny. Similar to the WGB genome, that of WGM has retained the capacity to synthesize functional flagella. To better understand the biological role of flagella, we quantified transcripts for flagellin (*fliC*), which encodes the filament subunit of bacterial flagella, and motility protein A (*motA*), which confers motility functions on flagella, using quantitative reverse transcription-PCR (qRT-PCR) and immunohistochemistry approaches (Fig. 4). We quantified gene expression in the maternal gut bacteriome organ, in different stages of intrauterine larvae (L1 to L3) and in the corresponding mothers' carcasses representing milk glands, and in young (newly deposited) and old (prior to eclosion) pupae. Within the tsetse fly mother, *motA* and *fliC* were expressed only in the carcass, apparently by WGM bacteria that are extracellular in the milk gland organ (Fig. 4A). We also detected the expression of flagellar components in the intrauterine larvae carried by the mothers and in the young pupae immediately postdeposition (Fig. 4A). The expression levels of both flagellar genes were highest in the L1 stage of the intrauterine larvae and in the carcasses (milk glands) of the corresponding mothers. Flagellum-specific expression in larvae decreased during development and was lowest during pupal development. Neither *fliC* nor *motA* expression was detected in adult bacteriomes. Immunohistochemistry analysis with antibodies specific for WGM flagellin also confirmed the expression profile. No flagellin was detected in the intracellular WGM in the maternal gut bacteriome, whereas flagellin was observed in milk gland

cells and in the newly formed bacteriome organ in the intrauterine larva (Fig. 4B).

Functional biology of *Wigglesworthia*. To understand the regulation and functions of *Wigglesworthia* genes during different host developmental stages, we examined the expression profiles of two genes (*hemH* and *thiC*) associated with heme and thiamine biosynthesis; respectively, which may be involved in host nutrient supplementation. We also evaluated the expression of *groEL*, which encodes a chaperonin that may compensate for the higher-frequency protein misfoldings typically associated with an accelerated mutation rate (Fig. 5). Interestingly, all of the genes except *groEL* exhibited tissue- and host development-specific transcriptional regulation. The *thiC* and *hemH* genes showed similarities in their transcriptional regulation, and their expression was highest during the pupal stage of host development. However, *thiC* expression was higher than *hemH* expression in the adult bacteriome organ (intracellular stage). In contrast, during intrauterine larval development (L1 and L2), the *hemH* level was significantly higher than that in the adult bacteriome. Interestingly, the *hemH* levels in the different larval stages (L1 to L3) were similar to those observed in the corresponding milk gland samples obtained from the mother (MomL1 to MomL3; Fig. 5). The chaperonin encoded by *groEL* was expressed more consistently throughout all of the host stages examined, presumably due to its required assistance in protein folding throughout host development. Thus, the symbiont genes analyzed were subject to spatial and temporal transcriptional regulation during host development.

Molecular evolution. Rates of synonymous (dS) and nonsynonymous (dN) nucleotide changes in genes common to WGM and WGB were estimated to identify potential targets of selection. Both the dN and dS methods (dnds and dndsml) estimated com-
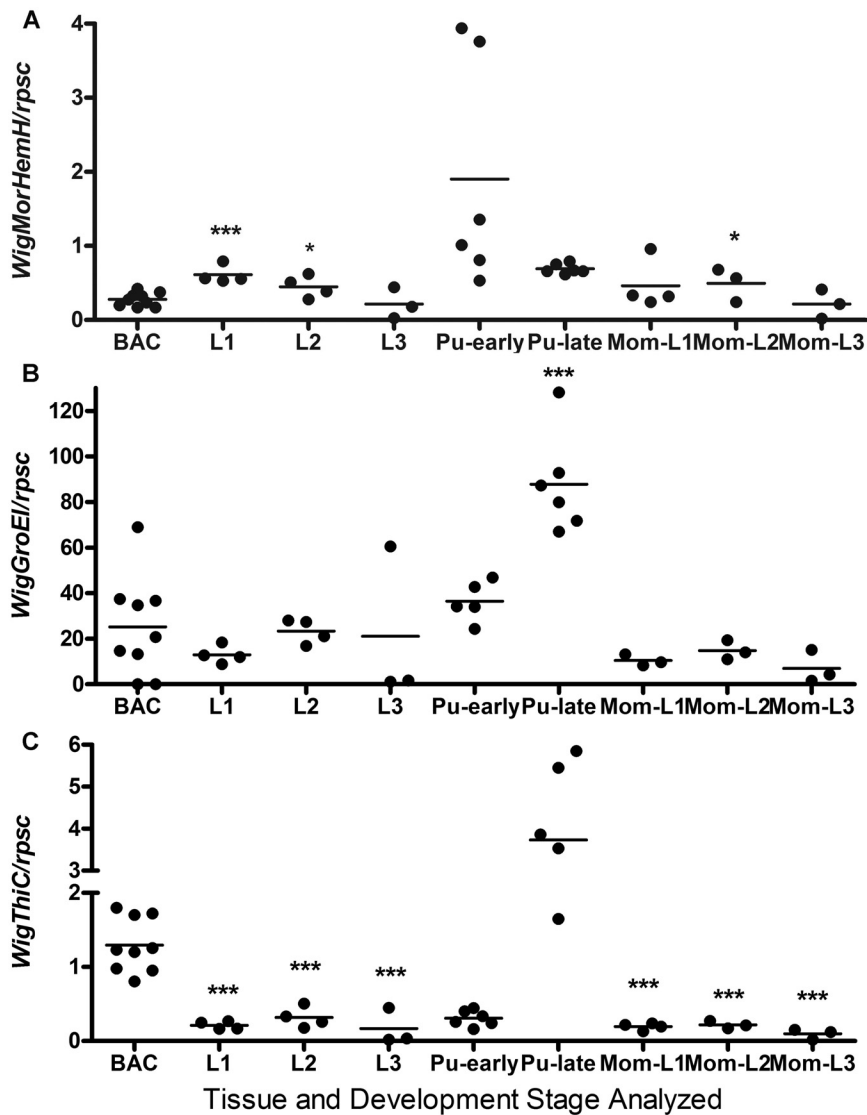
FIG 5 qRT-PCR of *hemH*, *groEL*, and *thiC* expression during tsetse fly development. Gene expression results are shown for *hemH* (A), *groEL* (B), and *thiC* (C) in the adult bacteriome (BAC), different stages of intrauterine larvae (L1, L2, and L3), newly deposited pupae (Pu-early), late pupae (Pu-late), and carcasses of mothers that carry the corresponding intrauterine larvae (Mom-L1, Mom-L2, and Mom-L3). All data were normalized to the ribosomal gene *rpsC* Asterisks indicate statistically significant differences between the bacteriome and different developmental stages. ***, $P < 0.0001$; *, $P < 0.05$.

parable values (see Table S3 in the supplemental material), so the data were discussed without the application of a maximum-likelihood-based correction. Rather than a direct inference of positive selection, we identified genes that have a higher rate of non-synonymous change than the rest of the genome.

Twenty-one gene comparisons were excluded from the dN and dS analysis due to length differences of >100 bp (see Table S3 in the supplemental material). Only one of these genes was detected in the six plasmid orthologs. Of these 21 genes, 7 had dN and dS values that were greater than 2 standard deviations above the mean, and most of these genes had large deletions, suggesting that these loci are not under positive selection but are undergoing degradation. Potential targets of selection are summarized in Table 2. Only two genes (*cspE* and *acpP*) were identified as likely targets of

purifying selection, although a few additional genes had relatively lower dN and dS values (Table 2; Fig. 6; see Table S3 in the supplemental material). Cold shock proteins (*cspE*) are associated with the maintenance of cellular function in cold temperatures and have been observed to function under osmotic stress (26). Acyl carrier proteins (*acpP*) are involved in cellular metabolism, particularly in fatty acid synthesis (27). Twenty-one genes were found to have significantly high values of dN and dS relative to the remainder of the genome (mean dN and dS = 0.2617), although only a single gene (*fliK*) was found to have a dN and dS value of >1, suggesting the influence of positive selection on this gene (Table 2). Importantly, genes that are potentially targets of selection were not restricted to a single area of the genome (Fig. 6) and one of these genes was found in the plasmid (WgpWb003). The functions of the genes varied, but notably, three genes with relatively higher dN and dS values that were involved in flagellar biosynthesis (*flgA*, *fliM*, *fliK*) and six cell surface-associated genes (*ppiD*, *yraP*, *ompF*, *yfiO*, *tolA*, and *imp*) were included. The plasmid gene WgpWb003 has homology to the *traT* gene that encodes a highly cell surface-exposed lipoprotein specified by F-like plasmids known to impede the conjugative transfer of similar or identical plasmids (28). These loci encode proteins that are exposed to the host environment, and thus, the higher dN and dS values may indicate selection for varying host immunological backgrounds.

DISCUSSION

Despite almost 80 million years of evolutionary distance, comparative analyses of the WGM and WGB genomes reveal similarly reduced size, almost complete synteny with the exception of one large inversion event, and a large set of genes shared by the two symbiont species. This is largely analogous to what has been described for other obligate insect endosymbionts (20, 23). This genome evolutionary process noted in obligate symbionts is distinct from what has been observed for free-living microbes, where significant diversity is driven by horizontal gene transfer on a background of gradual genome sequence drift. Despite high conservation, the two *Wigglesworthia* genomes display several unique capabilities, which are indicative of an adaptive evolutionary process. In particular, the putative proteomes indicate various metabolic capabilities in chorismate, phenylalanine, and folate biosynthesis, which in turn may affect their host physiology, including host vector competence (ability to transmit pathogenic trypanosomes). Furthermore, distinct from other obligate mutu-

TABLE 2 Summary of genes found to have notable dN/dS ratios

Gene[a]	p distance[d]	Genome position	G+C content[e]	dN/dS ratio	Function
cspE[b]	0.0931	632650	0.3286	0.0000	Cold shock protein
acpP[b]	0.0109	101980	0.3376	0.1325	Acyl carrier protein
ppiD	0.3635	662377	0.1675	0.5300	Outer membrane protein folding
tsf	0.3235	384266	0.1907	0.5335	Protein chain elongation
flgA	0.4000	45952	0.2153	0.5377	Flagellar biosynthesis
yrbK	0.3957	454967	0.1657	0.5387	Unknown
cyoD	0.3519	665149	0.1545	0.5519	Electron carrier
rrmJ	0.2965	568641	0.2259	0.5526	Methyltransferase
recC	0.3888	530384	0.1648	0.5547	DNA helicase
fliM	0.3302	58850	0.1729	0.5598	Flagellar energizing component
yraP	0.3909	311493	0.1968	0.5657	Lipoprotein
ygfZ	0.3791	327763	0.1849	0.5674	Folate-dependent regulation
yjgP	0.3389	304193	0.1685	0.5927	Transmembrane protein
folB	0.4453	297053	0.1861	0.5936	Dihydroneopterin aldolase component
ompF	0.4047	522497	0.1680	0.6185	Outer membrane porin
yfiO	0.3556	214973	0.1380	0.6389	Outer membrane protein component
holA[c]	0.4176	639753	0.1293	0.6892	DNA polymerase subunit
tolA[c]	0.3966	360872	0.1627	0.7200	Membrane-anchored protein
yeaZ[c]	0.4678	406494	0.1746	0.7509	Putative protease
tig[c]	0.4068	656418	0.1369	0.7526	Molecular chaperone
imp[c]	0.4183	20329	0.1418	0.8048	Envelope biosynthesis
fliK[c]	0.4699	60369	0.1383	1.0489	Flagellar hook protein
WgpWb003	0.3717	Plasmid	0.2069	0.5659	Transfer surface lipoprotein

[a] Genes without superscripts have dN/dS ratios that are ≥2 standard deviations above the mean.

[b] Gene with a notably low dN/dS ratio.

[c] Gene >3 standard deviations above the mean.

[d] p distance is the proportion of differences between the two species

[e] G+C content corresponds to that of the WGM gene copy.

alist symbiont systems, *Wigglesworthia* displays significant transcriptional regulation, including the expression of functional flagella in the extracellular forms present in mother's milk, which are apparently transmitted to host progeny. Significant levels of gene regulation at the transcriptional level have not been previously described in other ancient endosymbionts.

Eleven WGM genes demonstrated extensive (≥50%) truncation compared to their *E. coli* orthologs and were annotated as pseudogenes (see Table S1 in the supplemental material). Only three of these pseudogenes (*ftsK*, *nusB*, and *nlpB*) are similarly truncated in the WGB genome, suggesting ongoing, but relatively minor, gene degradation. Previously, Degnan et al. (23) questioned the pseudogene designation within the *Blochmannia* genome since the sequence conservation of affected genes suggests that they may still potentially encode functional proteins. Similarly, a majority of WGM pseudogenes (i.e., 6 out of 11) have frameshifts consisting of only 1 or 2 indels, including WGM *thiI*, which has previously been shown to be transcribed within adult bacteriomes (18). Various molecular processes occurring during transcription or translation may restore protein function in asso-
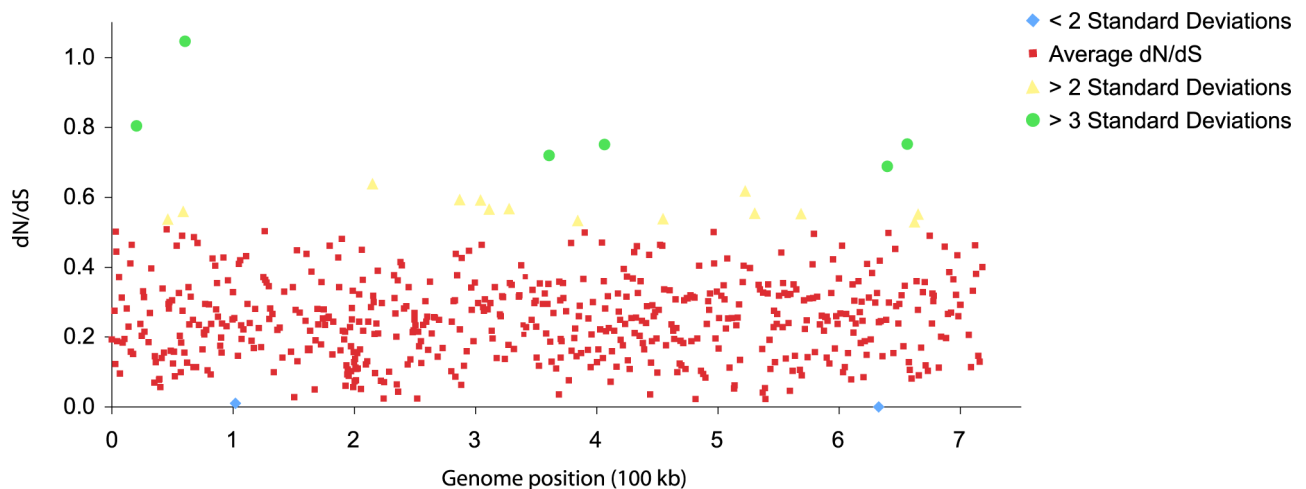

FIG 6 Summary of dN/dS ratio calculations relative to genome position. Genes putatively influenced by purifying selection are represented by blue diamonds. Genes within 2 standard deviations of the mean dN/dS ratios are represented by red squares. Genes with >2 and >3 standard deviations from the mean dN/dS ratio are represented by yellow triangles and green circles, respectively.

ciation with frameshift mutations, particularly within homopolymeric tracts (29). It is possible that these highly reduced genomes, by circumventing minor frameshift mutations, have evolved novel mechanisms to overcome the limitations of strict intracellular life.

Unlike the complete conservation in gene order and strand orientation reported within many ancient endosymbionts (20, 22, 23), a chromosomal inversion has occurred since the divergence of the WGM and WGB genomes. However, within the inversion, gene order has been retained between WGM and WGB. Recently, a smaller (~19-kb) inversion has also been described in the cockroach endosymbionts *Blattabacterium* (30) and a small (~7-kb) region within the *Tremblaya princeps* genome has been found in both orientations within the mealybug host populations (31). Nearly identical plasmid complements are harbored within WGB and WGM cells. The similar G+C contents of pWgm and its resident genome suggest early acquisition followed by a lengthy coevolution with the *Wigglesworthia* lineage. Furthermore, the uniformity between pWgm and pWgb in size, G+C content, and gene content and order suggests exposure to similar evolutionary processes over time. The stasis of the *Wigglesworthia* plasmids, relative to gene content and order, is in contrast to the versatility reported for the *Buchnera* extrachromosomal elements (32, 33) and may be attributable to particularities of insect host species ecology. The retention of these genes and genome elements by both the WGM and WGB genomes suggests their importance in *Wigglesworthia* biology and the symbiosis within the tsetse fly host background.

Only a small set of genes occurs in only one of the *Wigglesworthia* genomes (i.e., unique genes). These genes are either absent or still identifiable as a pseudogene in the sister genome (Fig. 3B; see Table S2 in the supplemental material). The retention of these unique genes may provide insight into the functional adaptation and evolution of the endosymbionts following tsetse fly host divergence. However, the genomes of the obligate symbionts investigated to date have undergone drastic size reduction and appear to continuously lose genes due to random reductive processes (2, 4). Thus, the suite of unique genes in each of the *Wigglesworthia* genomes may reflect remnants of these random processes rather than species-specific interactions. Despite this, the presence of constituents of genetic pathways, which are widely dispersed throughout the host chromosome, does argue for selection favoring the retention of these loci. Whether these genes encode functional pathways and how they factor into host biology and ecology remain to be examined.

Interestingly, although the WGB and WGM genomes retain similar numbers of unique genes (Fig. 3B), these genes span a variety of functional classes. When the unique genes were classified by functional relevance, categories such as information transfer, regulation, transport, and hypothetical were quantitatively comparable (see Table S2 in the supplemental material). In relation to DNA processing (information transfer), the two *Wigglesworthia* genomes demonstrated distinctions, particularly in light of recombination and repair. Exclusively encoded within WGB are *uvrD*, involved in nucleotide excision repair and methyl-directed mismatch repair, *recJ*, the single-stranded-DNA-specific exonuclease necessary for many recombination events (34), *yqgF*, a putative Holliday junction resolvase (35), and the nucleotide exchange factor, *grpE* (36). Meanwhile, *mutY*, which is involved in the correction of error-prone DNA synthesis due to oxidative

stress (37), is present in WGM. Whether these differences in DNA repair and recombination reflect particular advantages in different host environments is unknown. It is possible that the retention of a suite of recombination-related genes by the WGB genome, including *recA* found in *Wigglesworthia* and *Blattabacterium* spp., may have contributed to the chromosomal inversions noted in both species. The absence of the *recA* gene, in particular, in many ancient endosymbiont genomes has been suggested to contribute to the chromosomal stability noted by the absolute conservation of genome colinearity (20). Genetic loci associated with the stabilization and maturation of ribosomal subunits, b2511 and *rimM*, were also differentially retained within the WGB and WGM genomes, respectively.

In addition, some of the unique genes retained in each genome (*surA*, *ygcS*, *ftsL*, *bacA*, *brnQ*, and b2817 in WGB and *znuA* and *yfgL* in WGM) encode cell surface-associated proteins. Symbiont surface proteins have been shown to be pivotal in the homeostasis of host-microbe relations, suggesting a possible role for these proteins in host species adaptation processes (38, 39). Unlike *Buchnera*, which is enclosed in host-derived vacuoles in bacteriocytes, *Wigglesworthia* lies free within the host cell cytosol in the bacteriome organ and has an extracellular stage in the milk in the accessory glands. Thus, cell surface proteins may be particularly relevant in host-symbiont interactions for *Wigglesworthia* symbiosis. In support of their divergence, signatures of Darwinian positive selection mean of 0.65 ± 0.04 (standard error) were noted in several *Wigglesworthia* membrane-associated proteins (encoded by *ppiD*, *yraP*, *ompF*, *yfiO*, *tolA*, and *imp*).

Some aspects of the WGM genome suggest that WGM can perform novel functions compared to WGB (see Table S2 in the supplemental material). The WGM unique gene set includes *poxA* and *yjeK*, which are functionally coordinated in the posttranslational modification of elongation factor P (EF-P), which is involved in protein synthesis (40). In a recent survey where 725 bacterial genomes were analyzed, all possessed an *efp* gene but only 28% possessed both the *poxA* and *yjeK* genes. In other organisms, including WGB, EF-P may be modified by another pathway or the translation machinery may have been adapted to cope with the lack of EF-P modification (40). Analysis of the WGM specific gene set also reveals distinct metabolic capabilities such as the presence of a complete shikimate biosynthetic pathway. Chorismate is required for the synthesis of all aromatic amino acids, as well as other vitamins and cofactors (41). Unlike *E. coli* and *Salmonella enterica* Typhi, which utilize a type I 3-dehydroquinate dehydratase (i.e., *aroD*) as the third enzymatic reaction in the shikimate pathway, WGM encodes a type II 3-dehydroquinate dehydratase (*aroQ*) that is shorter but orthologous to genes found in *Helicobacter pylori*, *Yersinia pestis*, and *Mycobacterium tuberculosis* and in other symbiont genomes such as those of *Buchnera* and *Blochmannia*. Interestingly, these type II 3-dehydroquinate dehydratases are homologous to fungal catabolic 3-dehydroquinases (42). Given that WGB is associated with the most ancestral tsetse fly species (*G. brevipalpis*), it is likely that unique genes in WGM were lost in WGB following the divergence of the WGM host lineage, likely due to random gene loss. Alternatively, unique genes in WGM may have been acquired by lateral transfer following host speciation, but such events are thought to be negligible in the evolution of obligate endosymbionts due to their intracellular localization and reduced recombination rates (2). Thus, the origin of the shikimate biosynthetic pathway requires further investiga-

tion into its absence or presence within different *Wigglesworthia* species.

Downstream of the chorismate pathway, WGM also contains genes involved in folate and phenylalanine biosynthesis. Intriguingly, African trypanosomes, i.e., *Trypanosoma brucei brucei* (43), are unable to synthesize phenylalanine and folate yet encode transporters to salvage both from the host environment. Whether these genomic differences between WGM and WGB contribute to variation in chorismate, phenylalanine, and folate biosynthetic capabilities and are involved in the higher vector competency (44, 45) of *G. morsitans* warrants further investigation.

Both WGB and WGM are genetically capable of flagellar synthesis, with associated genes demonstrating dN and dS values approximately equivalent to those of the remainder of the genome (dN and dS average of 0.28 ± 0.04 compared to a genome-wide average of 0.2617 ± 0.005; see Table S3 in the supplemental material), suggesting the potential for selection to be acting to preserve genes of biological importance. Here we demonstrate that *fliC* and *motA*, which are associated with structural and motility functions, respectively, are specifically expressed at particular host life stages, notably, during the maternal transmission process and larval intrauterine development (Fig. 4). Moreover, hybridization of *Wigglesworthia*-specific FliC antibodies within the milk glands of gravid females and within the newly formed bacteriome organs of larval progeny further supports the role of flagella in *Wigglesworthia* transmission. Thus, it appears that flagella may play a role in both the transmission of *Wigglesworthia* from mother to progeny in milk and in the colonization of the larval bacteriome in the intrauterine larva early in development.

Regulation of genes at the transcriptional level has not been previously described in other ancient mutualistic endosymbionts, such as *Buchnera* (46, 47), and only very modest levels (i.e., rarely exceeding a factor of 3) have been described for *Blochmannia* (48). Although tissue-specific regulation of ankyrin domain-encoding genes by the parasitic endosymbiont *Wolbachia* within the gonads of multiple *Drosophila* spp. has been observed, these were also at comparably low levels (49). We examined the expression of two genes, *thiC* and *hemH*, associated with thiamine and heme biosynthesis; respectively, which are thought to be involved in host nutrient supplementation. Both *thiC* and *hemH* exhibited the highest levels of expression during the pupal stage of host development, a metabolically expensive period during insect metamorphosis, when adult morphological features develop with no food intake (Fig. 5). Interestingly, our prior symbiont density studies had indicated that the early pupal stage harbored relatively few *Wigglesworthia* cells, with marked proliferation occurring late in pupal development (50). Thus, it is tempting to speculate that the high metabolic demand on *Wigglesworthia* during the tsetse fly pupal stage may serve as a cue for its proliferation. A high level of *thiC* expression was also detected in intracellular *Wigglesworthia* in the adult bacteriome, supporting the significance of vitamin B supplementation in host nutrient provisioning. In contrast, *hemH* levels were significantly lower in the adult bacteriome organ than in other life stages. It is likely that in the midgut there is excess heme acquired through the blood diet, while *Wigglesworthia*-synthesized heme may help provision iron during intrauterine and pupal developmental stages. In contrast, the chaperonin encoded by *groEL* was expressed more consistently throughout all of the host stages examined, presumably due to its required assistance in protein folding. Our ongoing experiments where global

gene expression is being investigated from different developmental stages will shed further light on symbiont functional biology and transcriptional regulation, as well as host-symbiont dialogue.

Comparison of the WGM and WGB genomes indicates high levels of synteny and functional conservation. Despite this, similarity and variation in genome composition between WGB and WGM have allowed us to make and test specific hypotheses regarding the functional biology of *Wigglesworthia* (e.g., flagellar expression, nutritional supplementation) and form the basis of future experimental analyses. For example, a high dN in the genes of obligate symbionts may be indicative of genome degradation or diversifying selection. Examination of additional *Wigglesworthia* genomes will now shed light on these processes. Our expression studies with the *thiC*, *hemH*, *groEL*, *fliC*, and *motA* genes indicate significant levels of transcriptional regulation and development- and tissue-specific functional roles for the symbiosis previously not observed in other obligate symbionts. Genome-wide analyses of gene expression in different host developmental stages and tissues are needed to better understand host-symbiont cross talk. In addition to tsetse fly host nutrient provisioning, the presence of *Wigglesworthia* during larval development has been associated with host immune maturation. Based on comparative genome analysis, we speculate another possible role for *Wigglesworthia* symbiosis where infections with pathogenic trypanosomes may depend upon symbiont species-specific metabolic products and thus influence the vector competence traits of different tsetse fly host species.

MATERIALS AND METHODS

Insects and DNA preparation. Genomic DNA from *Wigglesworthia* sp. (WGM) harbored by the tsetse fly *G. m. morsitans* was prepared. The *G. m. morsitans* colony maintained in the insectary at Yale University was originally established from puparia originating from fly populations in Zimbabwe. Flies are maintained at 24 ± 1°C and 50 to 55% relative humidity and received defibrinated bovine blood every 48 h by an artificial-membrane system (51). The bacteriome organs were isolated from about 1,000 adult females by dissection, bacteriocytes were released by gentle homogenization of the tissue, and DNA was isolated as previously described (16).

Sequencing methodology. The genome sequence of WGM was determined by the whole-genome shotgun strategy using Sanger sequencing. Genomic DNA was amplified by multiple-displacement amplification using a REPLI-g Midi kit (Qiagen) to obtain a sufficient amount of DNA for sequencing. The amplified genomic DNA was sheared using a HydroShear (Gene Machine). DNA fragments were fractionated by agarose gel electrophoresis and subcloned into vector plasmid pTS1 (Nippon Gene) to construct a shotgun library with an average insert size of 3 kb for sequencing using a 3730xl sequencer (Applied Biosystems). Template DNA was prepared by PCR with Ex-Taq (Takara Bio) on an aliquot of the bacterial culture to amplify the insert DNA of each clone. We produced 9,984 reads by sequencing both ends of the clones, giving 9.4-fold coverage. The assembly generated 14 contigs. Gap closing and resequencing of low-quality regions in the assembled data were performed by PCR, primer walking, and direct sequencing of appropriate plasmid clones. The overall accuracy of the finished sequence was estimated to have an error rate of less than 1/10,000 bases (Phrap score of ≥40).

Genome annotation and alignment. The rapid annotation using subsystems technology (RAST) server (52) was used for automated gene prediction and annotation of the WGM genome sequence. Predictions of ortholog between WGM and E. *coli* K-12 strain MG1655 were performed using a BLASTP reciprocal best-hit analysis with a threshold cutoff of 30% amino acid identity and requiring at least 60% of both proteins in the alignment. Because the E. *coli* MG1655 genome has been manually cu-

rated, resulting in high-quality and more-comprehensive gene annotations than automated processes can generate, the product names and gene names were transferred to orthologous genes from WGM. These annotations, alongside the ones from RAST, are available through the ASAP database at http://asap.ahabs.wisc.edu/asap/home.php (53). Genome sequences of WGM and WGB were aligned using Mauve with the match seed weight parameter increased to 21, allowing for a more accurate alignment of AC-rich genomes (54, 55), and orthologous genes were extracted using the export ortholog function. The WGM genome sequence was circularly permuted based on the Mauve alignment to the corresponding start site from WGB.

Manual annotation. For this analysis, no ORF smaller than 50 codons was considered a gene. Each WGM and WGB CDS predicted to be unique based on Mauve alignment was manually reanalyzed based on the results of BLAST (56) and FASTA (57) sequence comparisons using the nonredundant database at NCBI. To determine if lineage-specific orthologs were present, nucleotide sequences of unique CDSs, flanking orthologs, and the intervening sequences from both WGB and WGM were manually examined by MacClade 4.08. Unique CDS nucleotides were then aligned with the intervening sequence and translated to inspect the amino acid alignment. All unique CDSs, relative to either the WGM or the WGB genome, have been classified into one of the class qualifiers based on hierarchical cellular functions of MultiFun (58) available in the ECOCYC database (http://www.ecocyc.org). Metabolic pathways were reconstructed using the reference pathways available for *E. coli* at EcoCyc and Kegg (59). CDSs proposed to be absent from WGM were similarly manually verified using WGB gene sequences. Manual nucleotide and amino acid sequence alignments were performed in MacClade 4.08. Sequences were determined to be orthologous (have shared ancestry) if nucleotide and amino acid sequences were similar and if start and stop codons were present in approximately the same position within the alignment.

Pseudogene annotation. Final manual inspection identified adjacent ORFs representing fragments of the same gene and truncated ORFs; therefore, CDSs less than half the length of their functional homologs in related species were categorized as pseudogenes. All of the pseudogenes identified in WGM and their functional classes are shown in Table S1 in the supplemental material.

Plasmid annotation. Identification of loci orthologous to the WGB plasmid was performed with BLAST at NCBI. A graphic display of the WGM plasmid map was generated using PlasMapper (60).

Molecular evolution. Orthologous CDSs were retrieved from WGM and WGB using alignment coordinates from Mauve. Manually annotated orthologous sequences were extracted by hand. Start and stop codons were removed from each pair of orthologs, and nucleotide sequences were then translated into amino acid sequences in MatLab. Using the Needleman-Wunsch algorithm (61), amino acid sequences were aligned with a gap penalty opening of 12 and a gap extension penalty of 4. The following calculations were made for each ortholog pair using the Bioinformatics Toolbox in MatLab: p distance, GC content (for each gene), dN/dS ratio, and maximum-likelihood-corrected dN/dS ratio. Once aligned by codons, sequences were converted back to nucleotides and the proportion of nucleotide differences (p distance) were calculated using the seqpdist function in MatLab. Alignments of sequences with large p distances (many nucleotide comparisons differed) were manually verified in MacClade.

To examine evidence of selection on genes or regions of the genome, the ratio of the number of nonsynonymous substitutions to the number of synonymous substitutions (dN/dS ratio) was calculated using the seqpdist, dnds, and dndsml functions in MatLab. The dndsml function incorporates a model of sequence evolution to minimally account for multiple hits in sequences. Typically, in comparisons of dN/dS ratios, low values suggest that genes are under purifying selection while high values are indicative of positive selection. This study is limited by the comparisons of genes between only two genomes, so evidence of selection does not explicitly incorporate the

organism's evolutionary history. Therefore, we instead examined the dN/dS ratio relative to the entire genome of each of the two organisms. Genes that were found to have a dN/dS ratio significantly higher or lower than the mean of the genome comparisons were identified as those that differed from the mean by 2 to 3 standard deviations. Symbiont genomes commonly are subjected to deletions and accumulation of nonsynonymous mutations that eventually lead to gene loss (5). As a result, examination of the dN/dS ratios of genes that have substantial deletions and are potentially no longer functional may also accumulate large numbers of nonsynonymous mutations and therefore generate false positives in our survey of dN/dS ratios. Therefore, genes that differed between the two genomes by >100 bp were excluded from final dN/dS ratio presentation (see Table S3 in the supplemental material).

FliC immunostaining and microscopy. Antibodies were generated against *E. coli*-expressed, 6×His-tagged recombinant FliC protein. Primers were designed to amplify the coding region from bp 591 to 1001 of the WGB *fliC* gene. Primer design included restriction sites to facilitate directional, in-frame cloning into the pET-28a 6×His tag expression vector (Novagen, Madison, WI) (primer sequences: *WgbFliC* forward, 5′-AGC ATGAGCTCGGAATTGAAATAAAAAGCACA; *WgbFliC* reverse, 5′-AG CATCTCGAGGATCCATTGTTAAAAACATTGAAA). The pET-28a—*WgbFliC* constructs were transformed into *E. coli* BL21, and recombinant protein expression was induced by treatment of cultures with 100 μM isopropyl-β-d-thiogalactopyranoside. Bacteria were lysed by sonication, and products were analyzed by SDS-PAGE. RecFliC was found predominantly in the insoluble fraction as inclusion bodies. Inclusion bodies were solubilized in binding buffer in the presence of 6 M urea and purified by using nickel resin under a denaturing conditions protocol (Novagen His-Bind kit). RecFliC proteins were subsequently purified by SDS-PAGE, and gel slices were provided for commercial antiserum production (Cocalico Biologicals).

For immunohistochemistry, tissues were dissected and fixed for about 1 week in 4% paraformaldehyde. Samples were then dehydrated, embedded in paraffin, cut into 5-μm-thick sections, and mounted on poly-l-lysine-coated microscopy slides. After being dewaxed for 2 × 15 min in methylcyclohexane and 2 × 10 min in ethanol, samples were air dried and rehydrated in 1× phosphate buffer saline containing 0.01% Tween 20 (PBST). After 1 h of blocking in 3% bovine serum albumin in 1× PBST at room temperature, sections were incubated in WGB FliC antibody solution (1:500 in 1× PBST) overnight at 4°C. Following three 10-min washes in PBST, slides were incubated in anti-rabbit Alexa 488 antibody (Molecular Probes; diluted 1:500 in PBST) for 1 h at room temperature in the dark. Sections were then washed again three times for 10 min each time in PBST, rinsed in water, and air dried in the dark. They were then mounted in GelMount mounting medium, which contained 4′,6-diamidino-2-phenylindole (DAPI) and covered with coverslips. Microscopic analyses were conducted using a Zeiss Axioskop 2 microscope equipped with an Infinity 1 USB 2.0 camera and software (Lumenera Corporation). Fluorescent images were taken using a fluorescence filter set with fluorescein- and DAPI-specific channels

Real-time qRT-PCR gene expression analyses. For *Wigglesworthia* gene expression, gene-specific primers were used to quantify *hemH*, *motA*, *groEL*, *fliC*, and *thiC* transcripts. The *rpsC* gene was used for normalization. qRT-PCR was performed with an iCycler iQ real-time PCR detection system (Bio-Rad, Hercules, CA) using the primer sets and conditions described in Table S1 in the supplemental material. The normality of sample means from each treatment was determined by Shapiro-Wilk test prior to *t* test analysis. Values are represented as the mean ± the standard error of the mean, and statistical significance was determined using a Student's *t* test and Microsoft Excel software.

Nucleotide sequence accession number. The sequence data obtained in this study have been deposited in GenBank under project accession no. CP003315.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00240-11/-/DCSupplemental.

Figure S1, PPT file, 0.1 MB.
Figure S2, PPT file, 0.2 MB.
Table S1, DOCX file, 0.1 MB.
Table S2, DOCX file, 0.1 MB.
Table S3, DOCX file, 0.1 MB.

REFERENCES

1. **Buchner P.** 1965. Endosymbiosis of animals with plant microorganisms, p. 210–338. Interscience Publishing Group Inc., New York, NY.
2. **Moran NA, McCutcheon JP, Nakabachi A.** 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. **42**:165–190.
3. **Clark M, Moran N, Baumann P.** 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol. Biol. Evol. **16**:1586–1598.
4. **Wernegreen JJ.** 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. **3**:850–861.
5. **Moran NA, McLaughlin HJ, Sorek R.** 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science **323**:379–382.
6. **Chen X, Li S, Aksoy S.** 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. **48**:49–58.
7. **Toh H, et al.** 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. **16**:149–156.
8. **Dale C, Maudlin I.** 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans* morsitans. Int. J. Syst. Bacteriol. **49**:267–275.
9. **Cheng Q, Ruel TD, Zhou W, et al.** 2000. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. Med. Vet. Entomol. **14**:44–50.
10. **O'Neill SL, Gooding RH, Aksoy S.** 1993. Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. Med. Vet. Entomol. **7**:377–383.
11. **Pais R, Lohs C, Wu Y, Wang J, Aksoy S.** 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Appl. Environ. Microbiol. **74**:5965–5974.
12. **Ma WC, Denlinger DL.** 1974. Secretory discharge and microflora of milk gland in tsetse flies. Nature **247**:301–303.
13. **Attardo GM, et al.** 2008. Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity. J. Insect Physiol. **54**:1236–1242.
14. **Nogge G.** 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82**:101–104.
15. **Weiss BL, Wang J, Aksoy S.** 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol. **9**:e1000619.
16. **Akman L, et al.** 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. **32**:402–407.
17. **Akman L, Aksoy S.** 2001. A novel application of gene arrays: *Escherichia coli* array provides insight into the biology of the obligate endosymbiont of tsetse flies. Proc. Natl. Acad. Sci. U. S. A. **98**:7546–7551.
18. **Snyder AK, et al.** 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc, Biol. Sci. **277**:2389–2897.
19. **Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H.** 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature **407**:81–86.
20. **Tamas I, et al.** 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science **296**:2376–2379.
21. **Gil R, et al.** 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc. Natl. Acad. Sci. U. S. A. **100**:9388–9393.
22. **Van Ham RC, et al.** 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. U. S. A. **100**:581–586.
23. **Degnan PH, Lazarus AB, Wernegreen JJ.** 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. **15**:1023–1033.
24. **Tracy RB, Kowalczykowski SC.** 1996. *In vitro* selection of preferred DNA pairing sequences by the *Escherichia coli* RecA protein. Genes Dev. **10**:1890–1903.
25. **Rajan R, Wisler JW, Bell CE.** 2006. Probing the DNA sequence specificity of *Escherichia coli* RECA protein. Nucleic Acids Res. **34**:2463–2471.
26. **Phadtare S, Inouye M, Severinov K.** 2002. The nucleic acid melting activity of *Escherichia coli* CspE is critical for transcription antitermination and cold acclimation of cells. J. Biol. Chem. **277**:7239–7245.
27. **Misra A, Surolia N, Surolia A.** 2009. Evolutionary significance of self-acylation property in acyl carrier proteins. I.U.B.M.B. Life **61**:853–859.
28. **Harrison JL, Taylor IM, Platt K, O'Connor CD.** 1992. Surface exclusion specificity of the TraT lipoprotein is determined by single alterations in a five-amino-acid region of the protein. Mol. Microbiol. **6**:2825–2832.
29. **Wernegreen JJ, Kauppinen SN, Degnan PH.** 2010. Slip into something more functional: selection maintains ancient frameshifts in homopolymeric sequences. Mol. Biol. Evol. **27**:833–839.
30. **Sabree ZL, Degnan PH, Moran NA.** 2010. Chromosome stability and gene loss in cockroach endosymbionts. Appl. Environ. Microbiol. **76**:4076–4079.
31. **McCutcheon JP, von Dohlen CD.** 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. CB **21**:1366–1372.
32. **Gil R, Sabater-Muñoz B, Perez-Brocal V, Silva FJ, Latorre A.** 2006. Plasmids in the aphid endosymbiont *Buchnera aphidicola* with the smallest genomes. A puzzling evolutionary story. Gene **370**:17–25.
33. **Latorre A, Gil R, Silva FJ, Moya A.** 2005. Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. Heredity **95**:339–347.
34. **Lovett ST, Kolodner RD.** 1989. Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **86**:2627–2631.
35. **Aravind L, Makarova KS, Koonin EV.** 2000. Survey and summary: Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res. **28**:3417–3432.
36. **Wu B, et al.** 1996. Structure-function analysis of the *Escherichia coli* GrpE heat shock protein. EMBO J. **15**:4806–4816.
37. **Michaels ML, et al.** 1990 MutY, an adenine glycosylase active on G-A mispairs, has homology to endonuclease III. Nucleic Acids Res. **18**:3841–3845.
38. **Weiss BL, Wu Y, Schwank JJ, Tolwinski NS, Aksoy S.** 2008. An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. Proc. Natl. Acad. Sci. U. S. A. **105**:15088–15093.
39. **Nyholm SV, Stewart JJ, Ruby EG, McFall-Ngai MJ.** 2009. Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. Environ. Microbiol. **11**:483–493.
40. **Bailly M, de Crécy-Lagard V.** 2010. Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. Biol. Direct **5**:3.
41. **Pittard AJ.** 1987. Biosynthesis of the aromatic amino acids, in *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, p. 368–394. American Society for Microbiology, Washington, DC.
42. **Garbe T, et al.** 1991. The *Mycobacterium tuberculosis* shikimate pathway genes: evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases. Mol. Gen. Genet. **228**:385–392.
43. **Berriman M, et al.** 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science **309**:416–422.

44. **Harley JM, Wilson AJ.** 1968. Comparison between *Glossina morsitans*, *G. pallidipes* and *G. fuscipes* as vectors of trypanosomes of the *Trypanosoma congolense* group: the proportions infected experimentally and the numbers of infective organisms extruded during feeding. Ann. Trop. Med. Parasitol. **62**:178–187.

45. **Moloo SK, Kutuza SB.** 1988. Comparative study on the susceptibility of different *Glossina* species to *Trypanosoma brucei brucei* infection. Trop. Med. Parasitol. **39**:211–213.

46. **Moran NA, Dunbar HE, Wilcox JL.** 2005. Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. J. Bacteriol. **187**:4229–4237.

47. **Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA.** 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Mol. Microbiol. **48**:1491–1500.

48. **Stoll S, Feldhaar H, Gross R.** 2009. Transcriptional profiling of the endosymbiont *Blochmannia floridanus* during different developmental stages of its holometabolous ant host. Environ. Microbiol. **11**:877–888.

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

50. **Rio RV, Wu YN, Filardo G, Aksoy S.** 2006. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. Proc. Biol. Sci. **273**:805–814.

51. **Moloo SK.** 1971. An artificial feeding technique for *Glossina*. Parasitology **63**:507–512.

52. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

53. **Glasner JD, et al.** 2006. ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Res. **34**:D41–D45.

54. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.

55. **Rissman AI, et al.** 2009. Reordering contigs of draft genomes using the mauve aligner. Bioinformatics **25**:2071–2073.

56. **Altschul SF, et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

57. **Pearson WR, Lipman DJ.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A. **85**:2444–2448.

58. **Serres MH, et al.** 2001. A functional update of the *Escherichia coli* K-12 genome. Genome Biol. **2**:RESEARCH0035.

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

60. **Dong X, et al.** 2004. PlasMapper: a web server for drawing and auto-annotating plasmid maps. Nucleic Acids Res. **32**:W660–W664.

61. **Needleman SB, Wunsch CD.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.

62. **Lai CY, Baumann L, Baumann P.** 1994. Amplification of trpEG: adaptation of Buchnera aphidicola to an endosymbiotic association with aphids. Proc. Natl. Acad. Sci. U.S.A. **91**:3819–3823.