University of South Bohemia in České Budějovice
Faculty of Science

Arsenophonus and *Sodalis* Symbionts in Louse Flies: an Analogy to the *Wigglesworthia* and *Sodalis* System in Tsetse Flies

RNDr. Thesis

Eva Šochová

České Budějovice 2016

Šochová, E., 2016. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. RNDr. Thesis, in English – 11 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Anotation:

This study provides a new comparative model of symbiosis in sheep ked to well-characterized tsetse fly symbiotic system. Distribution of symbiotic associates was revealed by transmission electron microscopy and their evolutionary origin was confirmed by reconstruction of their phylogenies. Using genomic data, it explores their metabolic contribution to host and sheds light on B-vitamin metabolism.

Prohlašuji, že svoji rigorózní práci jsem vypracovala samostatně, pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své rigorózní práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 22. dubna 2016

Eva Šochová

AEM
Journals.ASM.org

Arsenophonus and *Sodalis* Symbionts in Louse Flies: an Analogy to the *Wigglesworthia* and *Sodalis* System in Tsetse Flies

Eva Nováková,[a,b] Filip Husník,[a,b] Eva Šochová,[a] Václav Hypša[a,b]

Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic[a]; Institute of Parasitology, Biology Centre, ASCR, v.v.i., Ceske Budejovice, Czech Republic[b]

Symbiosis between insects and bacteria result in a variety of arrangements, genomic modifications, and metabolic interconnections. Here, we present genomic, phylogenetic, and morphological characteristics of a symbiotic system associated with *Melophagus ovinus*, a member of the blood-feeding family Hippoboscidae. The system comprises four unrelated bacteria representing different stages in symbiosis evolution, from typical obligate mutualists inhabiting bacteriomes to freely associated commensals and parasites. Interestingly, the whole system provides a remarkable analogy to the association between *Glossina* and its symbiotic bacteria. In both, the symbiotic systems are composed of an obligate symbiont and two facultative intracellular associates, *Sodalis* and *Wolbachia*. In addition, extracellular *Bartonella* resides in the gut of *Melophagus*. However, the phylogenetic origins of the two obligate mutualist symbionts differ. In *Glossina*, the mutualistic *Wigglesworthia* appears to be a relatively isolated symbiotic lineage, whereas in *Melophagus*, the obligate symbiont originated within the widely distributed *Arsenophonus* cluster. Although phylogenetically distant, the two obligate symbionts display several remarkably similar traits (e.g., transmission via the host's "milk glands" or similar pattern of genome reduction). To obtain better insight into the biology and possible role of the *M. ovinus* obligate symbiont, "*Candidatus* Arsenophonus melophagi," we performed several comparisons of its gene content based on assignments of the Cluster of Orthologous Genes (COG). Using this criterion, we show that within a set of 44 primary and secondary symbionts, "*Ca*. Arsenophonus melophagi" is most similar to *Wigglesworthia*. On the other hand, these two bacteria also display interesting differences, such as absence of flagellar genes in *Arsenophonus* and their presence in *Wigglesworthia*. This finding implies that a flagellum is not essential for bacterial transmission via milk glands.

Evolution of insect-bacterium symbiosis has resulted in a variety of associations in a broad range of insect taxa. Many traits of these associations (e.g., specifics of the arrangement of the host and symbiont metabolism, location of the symbionts, etc.) have been shown to reflect the ecological type of the symbiosis (i.e., obligate mutualism versus facultative symbiosis) and the host's nutritional demands due to the insufficiency of the diet (1–3). For example, phytophagous insects feeding on plant sap (e.g., aphids, whiteflies, or psyllids) and exclusively hematophagous insects (e.g., tsetse flies, sucking lice, and bedbugs) are the two most frequently studied ecological groups harboring obligate mutualistic symbionts (1, 3). However, compared to the large amount and complexity of the data accumulated on the sap-feeding insects, our knowledge on the symbioses in hematophagous insects is still limited, despite the number of species playing a crucial role as vectors for numerous pathogens, often causing a major burden to public health and world economies.

Blood feeding originated independently in several insects groups. Particularly, flies (Diptera) accommodate a large number of blood-feeding specialists. Among them, four families of the Hippoboscoidea group, i.e., Glossinidae (tsetse flies), Hippoboscidae (louse flies), Nycteribiidae, and Streblidae (bat flies) form a monophyletic cluster (4) and share another fundamental biological feature, unique among insects, adenotrophic viviparity (5). Larvae of these flies undergo their development in the female's uterus and are deposited shortly before pupation. Their nutrition is provided by the mother via accessory glands (called milk glands), tubular organs opening into the uterus (6).

In tsetse flies (Glossinidae), the blood-feeding strategy together with the viviparous reproductive mode gave rise to a multipartite symbiotic system. The two principal symbionts are the obligate mutualist *Wigglesworthia glossinidia*, housed intracellularly within a bacteriome organ adjacent to the anterior midgut (7), and a facultative symbiont, *Sodalis glossinidius*, present in almost all tissues (8). Both symbionts are vertically transmitted via milk glands (7, 9, 10) and are presumed to compensate for the nutritionally unbalanced blood diet (11–15). This highly specific microbiome of tsetse flies is usually accompanied by reproductive manipulators from the genus *Wolbachia* (16) and a diversity of other transient bacteria of unknown relationship to the host (17, 18). In contrast to Glossinidae, the family Hippoboscidae is a species-rich, highly diversified, and cosmopolitan group feeding on mammals and birds (4). Individual genera and species differ in their host specificities and life styles. While some resemble tsetse flies in being temporal parasites that seek the host only for short-term blood feeding, others shed their wings after finding the host,

Received 6 May 2015 **Accepted** 23 June 2015

Accepted manuscript posted online 6 July 2015

Citation Nováková E, Husník F, Šochová E, Hypša V. 2015. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. Appl Environ Microbiol 81:6189–6199. doi:10.1128/AEM.01487-15.

Editor: H. Goodrich-Blair

Address correspondence to Eva Nováková, novaeva@paru.cas.cz.

E.N. and F.H. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01487-15.

doi:10.1128/AEM.01487-15

and some retain only highly reduced wings or may even be permanently wingless.

During our previous research, we have shown that members of the family Hippoboscidae host several symbiotic bacteria (19–21). For instance we have identified *Sodalis* and *Arsenophonus* bacteria as common associates of Hippoboscidae, likely experiencing multiple origins throughout the host evolution (21). In the manuscript, we characterized the symbiotic system of the sheep ked, *Melophagus ovinus* (Hippoboscidae), a Palearctic wingless species that spends its entire life on the host and is transmitted only by physical contact between sheep (22). Living exclusively on vertebrate blood that lacks B vitamins, the sheep ked requires obligate symbiotic bacteria for its successful growth and/or reproduction. To provide insight into the overall arrangement of this symbiosis, we present results from transmission electron microscopy (TEM) and fluorescent *in situ* hybridization (FISH) and describe morphological properties of the symbiotic organ, the bacteriome, and other tissues harboring symbionts. Furthermore, using a shotgun metagenomic approach, we identify the principal symbionts, members of the genus *Arsenophonus* and *Sodalis*, and provide their phylogenetic and genomic characteristics. To fully understand interactions within the system, we also considered two additional bacteria present in the sheep keds, *Bartonella melophagi* (here referred to as *B. melophagi*) and *Wolbachia* sp. While we obtained genome assemblies for *Arsenophonus*, *Sodalis*, and *Bartonella* symbionts, *Wolbachia* raw data were of extremely low coverage (approximately 8-fold) and were thus not further analyzed.

Since we show that the symbiotic species "*Candidatus* Arsenophonus melophagi" and "*Candidatus* Sodalis melophagi" (here referred to as "*Ca*. Arsenophonus" and "*Ca*. Sodalis," respectively) have originated independently of the symbionts in tsetse flies, we suggest that their comparison to the symbionts in Glossinidae may help in distinguishing phylogenetic constraints, adaptive processes, and contingent traits. We also propose *M. ovinus* as a potential comparative model for further investigation into symbiosis of the blood-feeding insect groups with adenotrophic viviparity.

MATERIALS AND METHODS

Insect samples, dissections, light microscopy, and TEM. Samples of *M. ovinus* were obtained from a licensed family sheep farm at Krásetín, Czech Republic (48°53′4.549″N, 14°18′51.894″E). Bacteriomes and milk glands were dissected in phosphate-buffered saline (PBS) under an Olympus SZ51 dissecting microscope. The bacteriome tissue samples were used for both DNA extraction and fluorescent *in situ* hybridization (FISH). Transmission electron microscopy (TEM) of milk glands was carried out as reported previously for bacteriomes (19).

Sample preparation, probe design, and FISH. The tissue samples including gut and reproductive tract with surrounding milk glands were fixed overnight in 4% paraformaldehyde and then incubated in 2% H_2O_2 for up to 48 h to quench tissue autofluorescence. Afterwards, tissues were embedded in paraffin, and 6-μm-thick sections were placed on silane-coated slides. The slides were kept dry at a room temperature until used in fluorescent *in situ* hybridization (FISH). The probes were designed to match specifically the 16S rRNA of each associated bacteria to avoid any cross-hybridization. The labeled probe sequences are as follows: "*Ca*. Arsenophonus," Flc-GGCCTTACGGTCCCTCA (where Flc is fluorescein); "*Ca*. Sodalis," Cy3-CATCGCCTTCCTCCCCGCTG; *B. melophagi*, TxRd-CGTCATTATCTTCACCGGTG (where TxRd is Texas Red); *Wolbachia* sp., Cy5-CTTCTGTGAGTACCGTCATTATC and Cy3-AACGCTAGCC CTCTCCGTA; these allow various probe combinations in a single hybridization. The sections were dewaxed in xylene for 10 min, followed by 10 min in a xylene-ethanol equal-ratio solution. After rehydration in a graded ethanol series, the samples were treated with a drop of 70% acetic acid and incubated at 45°C until dry. The samples were then dehydrated through a graded ethanol series and underwent deproteinization with proteinase K solution (Qiagen) for 8 min at 45°C. The slides were rinsed in PBS, dehydrated, and air dried prior to the prehybridization step, which was carried out for 30 min at 45°C using hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl, pH 7.2, 0.1% SDS). Hybridization was performed for 3 h at 45°C directly on slides coated with 200 μl of hybridization buffer with 2 ng/μl of specific, labeled probe for each of the bacterial associates. Afterwards, the slides were rinsed twice for 5 min in the hybridization buffer and briefly rinsed in PBS and water. The slides were mounted in Vectashield medium (Vector Laboratories) containing 4′,6-diamidino-2-phenylindole (DAPI) and kept in the dark at 4°C until observed on an Olympus FV1000 confocal microscope. Negative controls (using RNase-treated slides) and controls for probe specificity (employing a general eubacterial probe EUB338, Flc-GCTGCCTCCCGTAGGA) (23) were performed along with each hybridization (data not shown).

DNA isolation and PCR. Total genomic DNA (gDNA) for construction of an Illumina paired-end (PE) library was obtained from a single bacteriome. DNA was extracted using a QIAamp DNA microkit (Qiagen) according to the standard kit protocol. DNA concentration was determined using Qubit Fluorometric Quantification (Life Technologies), and its quality was assessed in 1% agarose gel using standard electrophoresis. In total, 3 μg of gDNA was used for PE library construction.

DNA isolates acquired from whole individuals were examined for "*Ca*. Sodalis," *B. melophagi*, and *Wolbachia* presence with diagnostic PCR and further used for *Wolbachia* multilocus sequence typing (MLST). MLST genes from *Wolbachia* were sequenced according to a *Wolbachia* MLST protocol (24). Diagnostic PCR with specific primers was performed on 82 flies (55 males and 27 females). Details on primer sequences, either newly designed or retrieved from literature (25, 26), and amplicon lengths are presented in Table S5 in the supplemental material. In order to standardize the PCR conditions, ready-to-use PPT master mix (Top Bio) was employed, and the sensitivity of PCR was evaluated on five different DNA concentrations ranging from 5 to 60 ng/μl. Products of several diagnostic PCRs were Sanger sequenced to ensure primer specificity.

"*Ca*. Sodalis" gDNA for Illumina 3-kb mate-pair (MP) library construction was extracted from a single colony cultivation of "*Ca*. Sodalis" spread on Mitsuhashi and Maramorosch insect (MMI) medium blood plate (19) using the same procedure. In total, 10 μg of gDNA was used for MP library construction.

Genome assemblies. Bacteriome gDNA was sequenced using an Illumina PE library with an average insert size of 180 bp. The PE and MP libraries were sequenced bidirectionally on one lane of a HiSeq 2000 system on 100-bp and 75-bp runs, respectively (Keck Microarray Resource, West Haven, CT, USA). In total, over 80 million reads were obtained from the PE library. Reads underwent adaptor and quality trimming, and those shorter than 25 bp were discarded. Processed reads were assembled as metadata using two different approaches. First, *de novo* assembly of all the processed reads into contigs was achieved on a CLC Genomic Workbench (CLC bio A/S) with default parameter settings. Second, a subset of roughly 50 million reads consisting of paired reads only was randomly reduced in order to adjust the average coverage for *Arsenophonus* data to 100-fold and assembled using Velvet (27). *De novo* assembly in Velvet was performed with a *k*-mer size set to 55 bp. Contigs from each assembly were binned separately based on their average coverage and blastx (28) hits against a nonredundant (nr) database. Two sets of almost identical *Arsenophonus* binned data were then compared and combined into the draft genome of 40 *Arsenophonus* contigs. The contigs were submitted to an automatic annotation pipeline of the Joint Genome Institute, Integrated Microbial Genomes/Expert Review (IMG/ER) (29), using Prodigal (30) as the preferred gene calling method.

A draft genome of "*Ca*. Sodalis" was assembled from a combination of PE and MP Illumina data by the Velvet assembler (27), with all parameters

determined by VelvetOptimiser (http://www.vicbioinformatics.com/software.velvetoptimiser.shtml) and default scaffolding. Automatic annotation for comparative purposes was executed in RAST (http://rast.nmpdr.org). A *B. melophagi* draft genome was assembled from PE data by following the same procedure. A *k*-mer size of 71 was used for *B. melophagi*, and a *k*-mer size of 57 was used for "*Ca*. Sodalis" because a 75-bp size was used for MP reads. The "*Ca*. Sodalis" genome was further scaffolded by SSPACE (31). Gaps in *B. melophagi* scaffolds were closed by PCR. The completeness of gammaproteobacterial draft genomes was evaluated according to the number of single-copy (SICO) genes (32) present in the assemblies assessed by blastp and tblastn searches (28).

Genome comparisons. Initial genome comparisons between "*Ca*. Arsenophonus" and *Wigglesworthia glossinidia* and between "*Ca*. Sodalis" and *S. glossinidius* were based on the elementary genome characteristics, e.g., genome size, GC content, number of coding sequences (CDS), and tRNA genes. For the more detailed analysis, orthologous gene clusters were used as the comparative data. This comparison, performed in a broader phylogenetic spectrum, was based on gene numbers within functional ortholog assignments as defined by the Cluster of Orthologous Genes (COG) database and implemented in the web project IMG (29). The data set was composed of gene numbers for 42 bacterial genomes retrieved by the abundance profile function in IMG/ER (29) and supplemented with the data from the genomes of the *Wolbachia* endosymbiont of *Cimex lectularius*, *Wolbachia* wCle, and "*Ca*. Sodalis" missing in IMG/ER (see Table S1 in the supplemental material). For the two latter genomes, the COG assignment was performed using the COG database (ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd) and reverse position-specific (RPS)-BLAST, which allows for comparison of a query against a library of position-specific scoring matrices (PSSMs) (33). Altogether, 2,473 orthologous clusters were included in the comparison. The analysis was based on two different measures, one expressed as the number of the genes present in the genomes and the other as a percentage of these genes calculated from the sum of all COG clusters. A dissimilarity matrix was calculated from these data using Bray-Curtis distance in R (http://www.R-project.org). Nonmetric multidimensional scaling ordination (NMDS) was used to show similarity in genome composition of selected symbiotic bacteria. Differences in overall genome composition, based on COG clusters, among different host feeding strategies, and between the character of symbiotic relationships (primary versus secondary) were assessed using permutational multivariate test implemented in the Adonis function in R (http://www.R-project.org).

Multiple alignments and phylogenetic analyses. Matrices for individual bacterial genera and their outgroups were compiled from the data obtained in this study and the data retrieved from public databases. Complete data sets were aligned using the Mafft algorithm (34) implemented in SeaView (35) software and concatenated into multigene matrices. These contained sequences for the *ftsK*, *fbaA*, *yaeT*, and *spoT* genes of *Arsenophonus*, the *gltA*, *groEL*, *ribC*, and *rpoB* of *Bartonella*, and the *gatB*, *coxA*, *hcpA*, *ftsZ*, and *fbpA* of *Wolbachia*. Individual concatenates were processed in the GBlocks application (36) in order to remove unreliably aligned positions. For the genus *Arsenophonus*, two matrixes were generated: the multigene concatenate described above and, in order to include "*Candidatus* Riesia pediculicola" into the analyses, an incomplete matrix containing the *fbaA* gene present in the extremely reduced genome of this bacterium. The resulting alignments were analyzed using maximum likelihood (ML) and Bayesian inference (BI). The best evolutionary models fitting the data were selected using the program jModelTest (37). ML-based analyses and 100 nonparametric bootstrap replicates were performed in the PhyML program (38). Bayesian analyses were performed with correspondent evolutionary models implemented in MrBayes, version 3.1.2 (39).

Reconstruction of metabolic pathways: B vitamins. Several independent approaches were employed to reliably reconstruct B-vitamin and cofactor pathways of all major bacteria present in *M. ovinus*-"*Ca*. Arsenophonus," "*Ca*. Sodalis," and *B. melophagi*. First, annotated draft genome assemblies were uploaded into Pathway Tools software using its pathological module (40), and individual pathways were compared with species comparisons and manually examined. Second, B-vitamin and cofactor genes from *Escherichia coli* K-12 MG1655 were searched by blastp (E value of 1e−5) against a custom database of all protein coding genes from all three bacteria. Third, putatively missing genes were checked by blasting homologues from *E. coli*, *S. glossinidius*, *Arsenophonus nasoniae*, and *Bartonella schoenbuchensis* species (blastx; E value, 1e−5) against a custom database of all contigs/scaffolds from bacteriome and "*Ca*. Sodalis" assemblies. B-vitamin and cofactor pathways were reconstructed with EcoCyc and KEGG databases as guides (41, 42).

Accession numbers. The "*Ca*. Arsenophonus" draft genome is provided in GenBank format (available at http://users.prf.jcu.cz/novake01/). The "*Ca*. Sodalis" and *B. melophagi* draft genomes are provided as fasta files with all scaffolds/contigs (http://users.prf.jcu.cz/novake01/). All genomes are currently in the process of gap closure and will be submitted to NCBI when closed into a single circular molecule (or several, in case of plasmids). Raw Illumina reads are available in the ENA database under accession number PRJEB9958.

RESULTS

Origin and distribution of the associated bacteria. Screening of the shotgun metagenomic data revealed five microorganisms in the sheep ked microbiota. Phylogenetic analyses showed that four of them belong to the well-known bacterial genera *Arsenophonus*, *Sodalis*, *Bartonella*, and *Wolbachia* (Fig. 1) and that one is the unicellular eukaryote *Trypanosoma melophagium* (see Fig. S1 in the supplemental material). The distribution of the bacteria in the host digestive and reproductive tracts, including overall arrangement of the bacteriome and likely transmission routes the symbionts undertake when passed to the progeny, is presented in Fig. 2 and 3.

The *Arsenophonus* bacterium possesses a number of features also found for some other obligate symbionts. The bacterium was present in all of the examined adults, housed in specialized cells of the host intestine wall (bacteriocytes), forming a compact symbiotic organ (bacteriome) (Fig. 2a to g). Under TEM, the bacteriocytes were tightly packed with pleomorphic populations of the symbiont cells (Fig. 2b). Apart from this typical location, in adult females, these bacteria also occur extracellularly in the lumen of the milk glands (Fig. 3a). This suggests that, similar to *Wigglesworthia* in tsetse flies, this *Arsenophonus* member also uses milk glands for its vertical transmission. Correspondingly, it was never observed in oocytes and early embryos, while its presence was confirmed in the late stages of embryogenesis (see Fig. S2 in the supplemental material). Based on the biological, phylogenetic, and evolutionary traits described in this study, we propose the designation "*Candidatus* Arsenophonus melophagi" (me-lo-phag-i; melophagi, of the genus *Melophagus*).

In contrast to the universally present "*Ca*. Arsenophonus," "*Candidatus* Sodalis melophagi" does not occur in all adult hosts (diagnostic PCR performed on 82 individuals from a single population indicates approximately a 50% prevalence, with 59% of females and 42% of males positive). In the gut, the bacterium resides predominantly outside the bacteriome. However, in contrast to *S. glossinidius*, which has never been described from the bacteriome, "*Ca*. Sodalis" may enter the symbiotic organ. Here, it occupies intercellular space surrounding individual bacteriocytes (Fig. 2e). Similar to "*Ca*. Arsenophonus," "*Ca*. Sodalis" was also detected in the milk glands. However, in addition to the extracellular location, it was also found intracellularly inside the secretory cells (Fig. 3a). Rarely, both bacterial symbionts co-occur in the

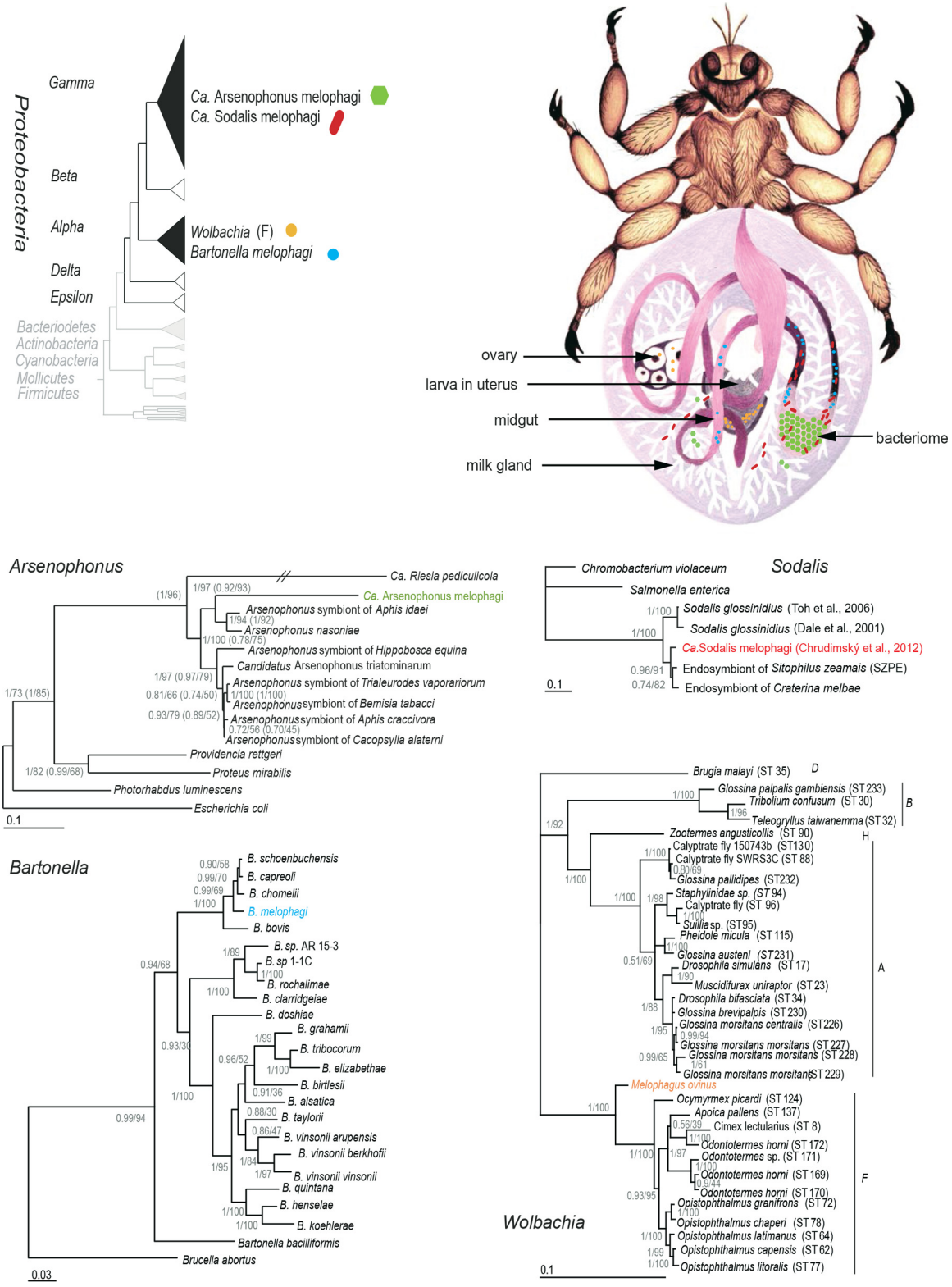

FIG 1 Phylogenetic origins and the overall arrangement of the *Melophagus* symbionts. Phylogenetic position and distribution of the bacteria associated with *Melophagus ovinus* are highlighted as follows: green, "*Ca.* Arsenophonus"; red, "*Ca.* Sodalis"; blue, *B. melophagi*; orange, *Wolbachia*. For *Wolbachia*, phylogeny branches are labeled with host names. Vertical bars and capital letters stand for *Wolbachia* supergroups. Numbers at branch nodes indicate posterior probability and bootstrap values. In *Arsenophonus* phylogeny, numbers in parentheses represent posterior probability and bootstrap values produced by the incomplete matrix with *fbaA* from "*Ca.* Riesia pediculicola" included. The references included are references 19 (Chrudimský et al., 2012), 67 (Dale et al., 2001), and 68 (Toh et al., 2006).

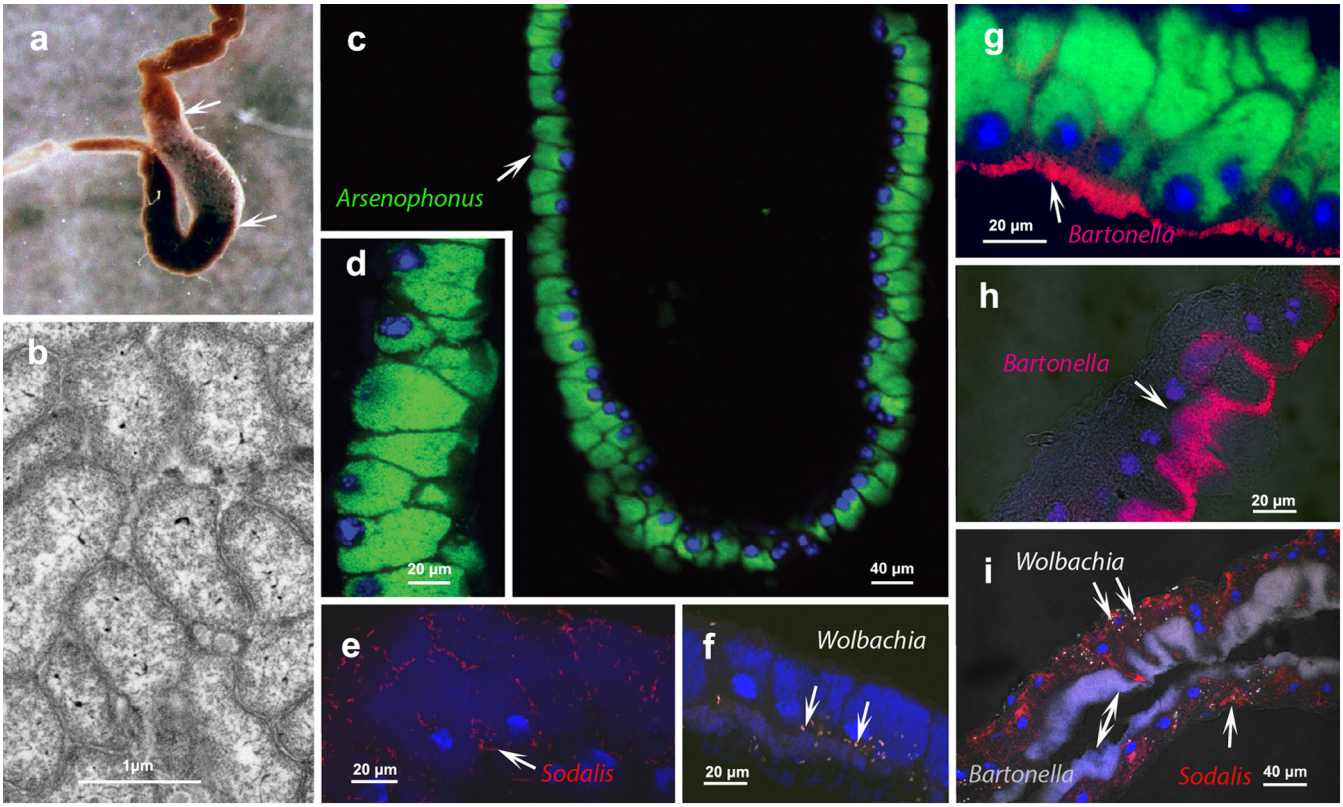

FIG 2 Light microscopy, fluorescence microscopy, and transmission electron microscopy (TEM) of *M. ovinus* gut sections. (a) Dissected midgut section with the symbiotic organ bacteriome designated with white arrows. (b) TEM of the bacteriome packed with "*Ca.* Arsenophonus" bacteria. Visualization Endosymbionts were visualized by 16S rRNA-targeted fluorescent *in situ* hybridization using multiple labeled probes: Flc, Cy3, Cy5, and Texas Red. Arrows and genus names indicate signals for different bacteria. (c, d, and g) "*Ca.* Arsenophonus" (Flc label) bacteria housed in the bacteriome. (e and f) Distribution of "*Ca.* Sodalis" (Cy3) and *Wolbachia* sp. (Cy5) throughout the bacteriome. *B. melophagi* (Texas Red) is present along the bacteriome (g) and in the rest of the midgut (h). (i) "*Ca.* Sodalis" (Cy3) and *Wolbachia* sp. (Cy5) in the midgut section. A nonspecific signal of *B. melophagi* (no probe) cells in the gut lumen is highlighted with a bidirectional arrow in panel i.

gland lumen (see Fig. S2 in the supplemental material). No "*Ca.* Sodalis" cells were detected in oocytes or early embryos.

Wolbachia found in *M. ovinus* branches as a sister lineage to the F supergroup (Fig. 1) and is highly prevalent in the screened population (100% of females and 94% of males positive) (see Materials and Methods for more detail on diagnostic PCR). *Wolbachia* was observed intracellularly in various tissues, including fat bodies (adipocytes) (Fig. 3b), secretory cells of the milk glands (Fig. 3a), and the gut tissue (Fig. 2f and i). Within the bacteriome, *Wolbachia* was scattered in small numbers among bacteriocytes (Fig. 2f). Although present in the secretory cells, *Wolbachia* was never observed in the milk gland lumen. In contrast to "*Ca.* Arsenophonus" and "*Ca.* Sodalis" symbionts, it infected embryos prior to intrauterine development (Fig. 3d). The findings described above distinguish the *Melophagus*-associated *Wolbachia* from the strains found in *Glossina*.

Bartonella melophagi was found in all of the screened individuals and is the only one of the detected bacteria located mostly extracellularly. It was found along the entire gut lumen (Fig. 2g and h) closely attached to the microvilli (see Fig. S1 in the supplemental material). Its abundance is particularly high along the gut wall outside the bacteriome section (Fig. 2h), whereas the cell numbers decrease within the bacteriome (Fig. 2g). Except for the gut, all the other tissues were found to be *Bartonella* free.

Genome comparisons. The shotgun metagenomic data combined with genomic data from "*Ca.* Sodalis" culture (see Materials and Methods) allowed for draft assembly of three bacterial genomes, i.e., "*Ca.* Arsenophonus," *Bartonella melophagi*, and "*Ca.* Sodalis," each with coverage above 100-fold (Table 1). The only bacterium with the coverage too low for further analysis was the symbiont of the genus *Wolbachia*. The sequencing and the assembly strategy we used resulted in genome assemblies truncated in the number of contigs/scaffolds by multiple rRNA operons ("*Ca.* Arsenophonus") and a number of duplicated genes and mobile genetic elements in the genomes of *Bartonella melophagi* and "*Ca.* Sodalis" (Table 1). The high coverage and the genome properties, particularly the number of tRNA genes and retained single-copy (SICO) genes in gammaproteobacterial genomes (see Materials and Methods), imply that the drafts represent the complete gene set. The *Bartonella* genome assembly fully reflects the gene content of the U.S. isolate of *B. melophagi* deposited in the NCBI Assembly database (GCA_000278255), sharing complete gene synteny and ~99% sequence identity in single-copy genome regions. Altogether, the basic genome characteristics listed in Table 2 underpin our microscopic findings, resembling the tsetse fly symbiosis. While the low-GC content and genome size of "*Ca.* Arsenophonus" are in line with its obligate relationship to the host, the con-

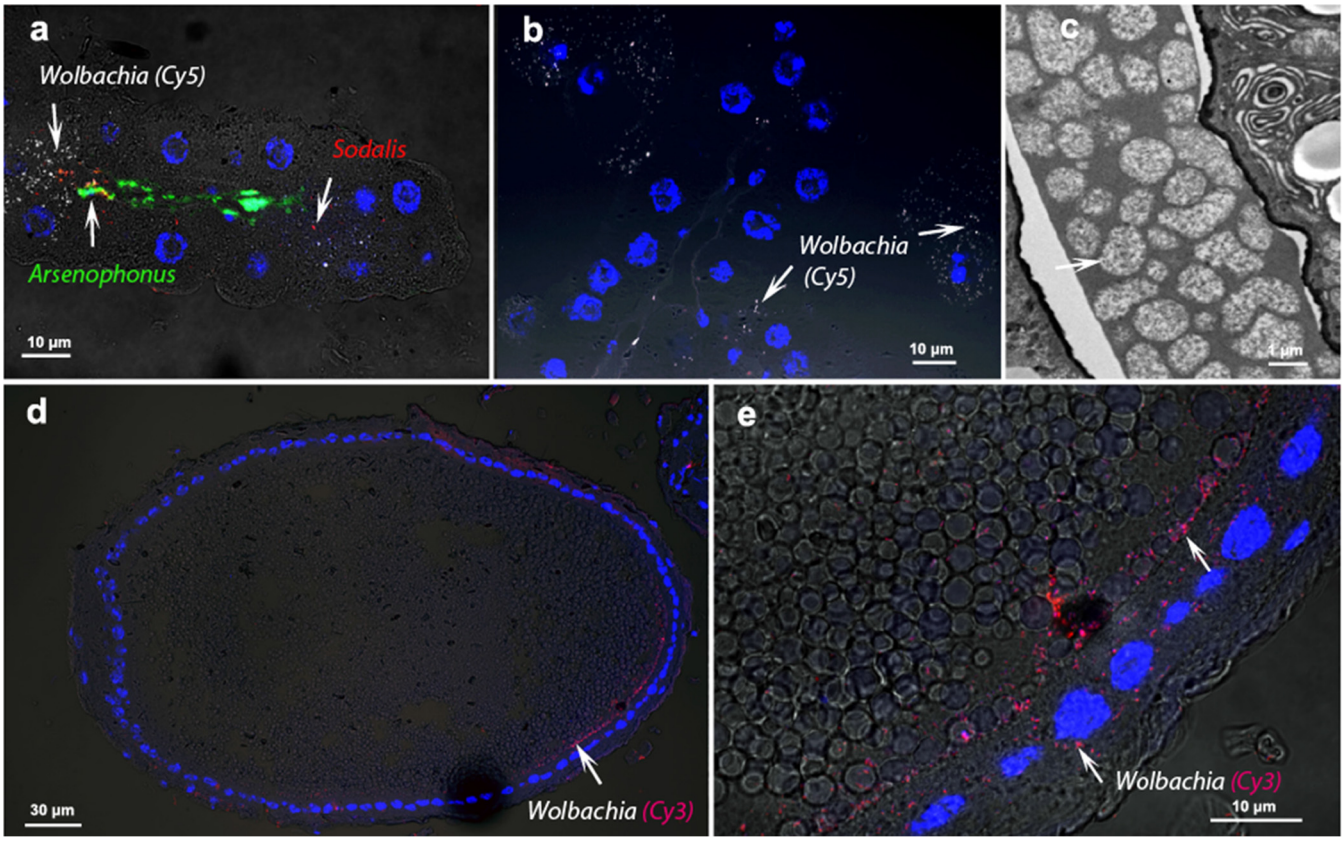

FIG 3 Fluorescence microscopy and transmission electron microscopy (TEM) of *M. ovinus* reproductive tract. Arrows and genus names indicate signals for different bacteria. (a) "*Ca.* Arsenophonus" (Flc label) bacteria present in the milk gland lumen, "*Ca.* Sodalis" (Cy3) in the lumen and intracellularly in the secretory cells of the gland, and *Wolbachia* sp. (Cy5) in milk gland secretory cells. (b) *Wolbachia* sp. in milk gland secretory cells and adipocytes. (c) TEM of the milk gland lumen containing "*Ca.* Arsenophonus" bacteria. (d) *Wolbachia* sp. in developing embryo. (e) Enlarged view of panel d.

siderably larger genome of "*Ca.* Sodalis" and its unbiased base composition suggest a less intimate association.

Compared to other *Arsenophonus* species and the bacterial symbionts from other taxa (see Table S1 in the supplemental material), "*Ca.* Arsenophonus" displays a remarkable degree of similarity with the tsetse fly symbiont *W. glossinidia* in terms of its genome content (Table 2; see also Table S2 in the supplemental material). Both genomes contain similar gene numbers and show similar patterns of present/absent orthologous genes. The ordination plot of overall similarities among 44 symbiotic bacteria, based on the NMDS analysis (Fig. 4), shows remarkably close distance between "*Ca.* Arsenophonus" and both *Wigglesworthia* genomes. As expected, the statistical evaluation of the clustering pattern also showed significant differences ($P = 0.001$) between genome composition of primary and secondary symbionts and among symbionts of hosts with different feeding strategies (i.e., blood-feeding, sap-sucking, and omnivorous insects) (Fig. 4).

Using the assignment of the orthologous clusters to individual

TABLE 1 Genome assembly metrics

Parameter	"*Ca.* Arsenophonus"	*Bartonella melophagi*	"*Ca.* Sodalis"
Approximate contig coverage (no.)	1,360	151	138
Total no. of contigs (no. of scaffolds)	40	74 (67)	616 (236)
Total no. of bases (including unresolved nucleotides)[a]	1,155,312	1,480,425	4,569,254
Total no. of unresolved nucleotides	0	180	423,006
Number of contigs with >1,000 bp	36	40	251
Largest contig (scaffold [bp])	116,563	502,641	153,948 (387.476)
Scaffold N50 (bp)[b]	87,029	381,476	171,607
Scaffold N75 (bp)[b]	33,043	144,087	95,759
GC content (%)	32.2	36.9	50.8
No. of present SICO genes[c]	203	N/A	205

[a] Introduced by the assembly process.
[b] N50 and N75, 50% and 75% of the length of the genome, respectively, is contained in contigs of the specified sizes or larger.
[c] Total number, 205.

TABLE 2 Comparison of basic genome characteristics between symbionts of *Melophagus ovinus* and *Glossina morsitans*

Parameter	*Melophagus ovinus* symbiont		*Glossina morsitans* symbiont	
	"*Ca.* Arsenophonus melophagi"	"*Ca.* Sodalis melophagi"	*Wigglesworthia glossinidia*	*Sodalis glossinidius*
Genome size (Mb)[a]	1.16	4.15	0.72	4.29
GC content (%)	32.2	50.8	25.2	54.5
Coding density (%)	61.3	84.9	87.9	78.4
No. of predicted CDS	765	4,964	678	5,812
tRNA genes (no.)	34	66	35	71
Protein coding genes (no.)	725	4,889	635	5,634
Average CDS length (bp)	926	653	933	579
No. of shared COGs	462	1,453	462	1,453

[a] Excluding unresolved nucleotides introduced by the assembly process.

COG categories (as defined by the KEGG database), the ratios of shared/present/absent orthologs was found to differ considerably across functional categories (see Tables S2 and S3 in the supplemental material). While in some categories (e.g., translation, ribosomal structure, and biogenesis; intracellular trafficking, secretion, and vesicular transport; and nucleotide transport and metabolism) the three genomes share similar subsets of genes, in others they differ considerably. Such a case is best illustrated by the categories replication, recombination, and repair, energy production and conversion, and cell motility. The "*Ca.* Arsenophonus" genome prevails in the first two categories, whereas the genome of *W. glossinidia* possesses a higher number of the genes in the last category. The cell motility category provides an extreme example where *W. glossinidia* possesses 28 genes, but the entire category is missing in "*Ca.* Arsenophonus." A further assignment of the genes within each COG category to particular metabolic pathways shows that the differences between the two symbionts in the gene counts are often due to absence/presence of a metabolic path or its part, rather than to random loss of individual genes (see Table S4 in the supplemental material).

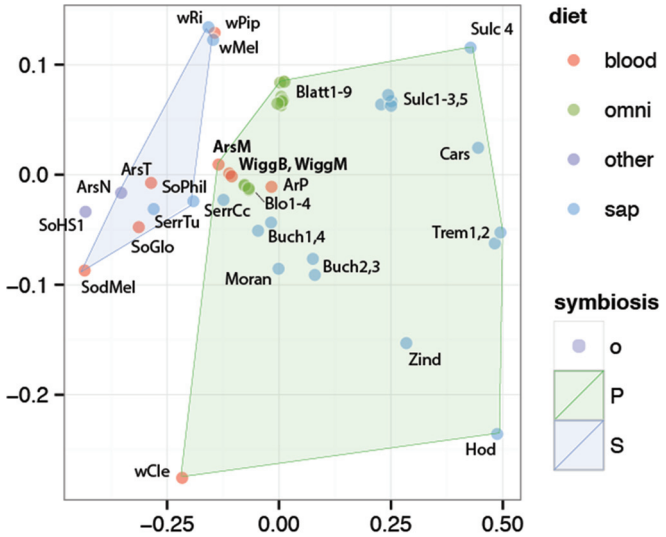

FIG 4 Genome-wide similarity among selected primary and secondary symbionts. The plot (constrained ordination) shows similarities among 44 bacterial genomes based on numbers of orthologous genes transformed into Bray-Curtis distances. ArM, "*Ca.* Arsenophonus melophagi"; WiggB and WiggM, *Wigglesworthia glossinidia* from *Glossina brevipalpis* and *Glossina morsitans*, respectively. A full list of the abbreviations and further information on analyzed genomes is provided in Table S1 in the supplemental material. P and S designate primary and secondary symbionts.

Metabolism: B-vitamin pathways. Because the crucial role of B vitamins produced by symbiotic bacteria in hematophagous systems has been generally accepted and because other possible functions of symbiotic bacteria in blood-feeding arthropods (such as immune system maturation) are impossible to infer based on genome data only, we streamlined the analyses toward B-vitamin biosynthetic pathways. Compared to the *Glossina*-associated symbiotic system, two of the B-vitamin pathways, those for thiamine and pantothenate, were arranged in notably different ways. The genes for the thiamine transporter are present in the obligate symbiont of *M. ovinus*, "*Ca.* Arsenophonus," while the facultative "*Ca.* Sodalis" symbiont codes for its complete biosynthetic pathway. *B. melophagi* is also capable of thiamine biosynthesis but uses an alternative pathway starting from glycine (Fig. 5). In correspondence with the absence of pantothenate biosynthesis genes (*panBCDE*) in the known *Arsenophonus* chromosomes (BioSample accession numbers SAMN02440620, SAMN02603451, and SAMN03078484), "*Ca.* Arsenophonus" also lacks these genes and carries only the genes necessary for synthesis of coenzyme A (CoA) from pantothenate (Fig. 5). On the other hand, the genome reduction in the obligate symbionts of blood-sucking insects led to losses of several identical enzymes in B-vitamin pathways, e.g., *phoA* in folate and *nadABC* in nicotinate pathways (see Fig. S3 in the supplemental material).

DISCUSSION

Biological convergence between the symbioses. The symbiotic system of *Melophagus ovinus* shows a remarkable overall analogy to that described for the related hippoboscoid group, the genus *Glossina*. Compared in more detail, these two systems possess an interesting combination of common features and unique characteristics. While the two obligate mutualists *W. glossinidia* and "*Ca.* Arsenophonus" originated independently from two phylogenetically distant groups, they share several important traits. The most apparent trait is their location in adult hosts within the bacteriomes formed by enlarged cells of the intestinal wall (Fig. 2). Considering their nutritional role, this seems to be a suitable location and has been reported for several other hosts, but it is certainly not the only and inevitable arrangement of symbionts in blood feeders, as documented by a variety of bacteriome structures and locations in sucking lice, bat flies, or bed bugs (43–45). Even more striking is the close resemblance of the transmission modes. Both bacteria, "*Ca.* Arsenophonus" and *W. glossinidia*, constitute two

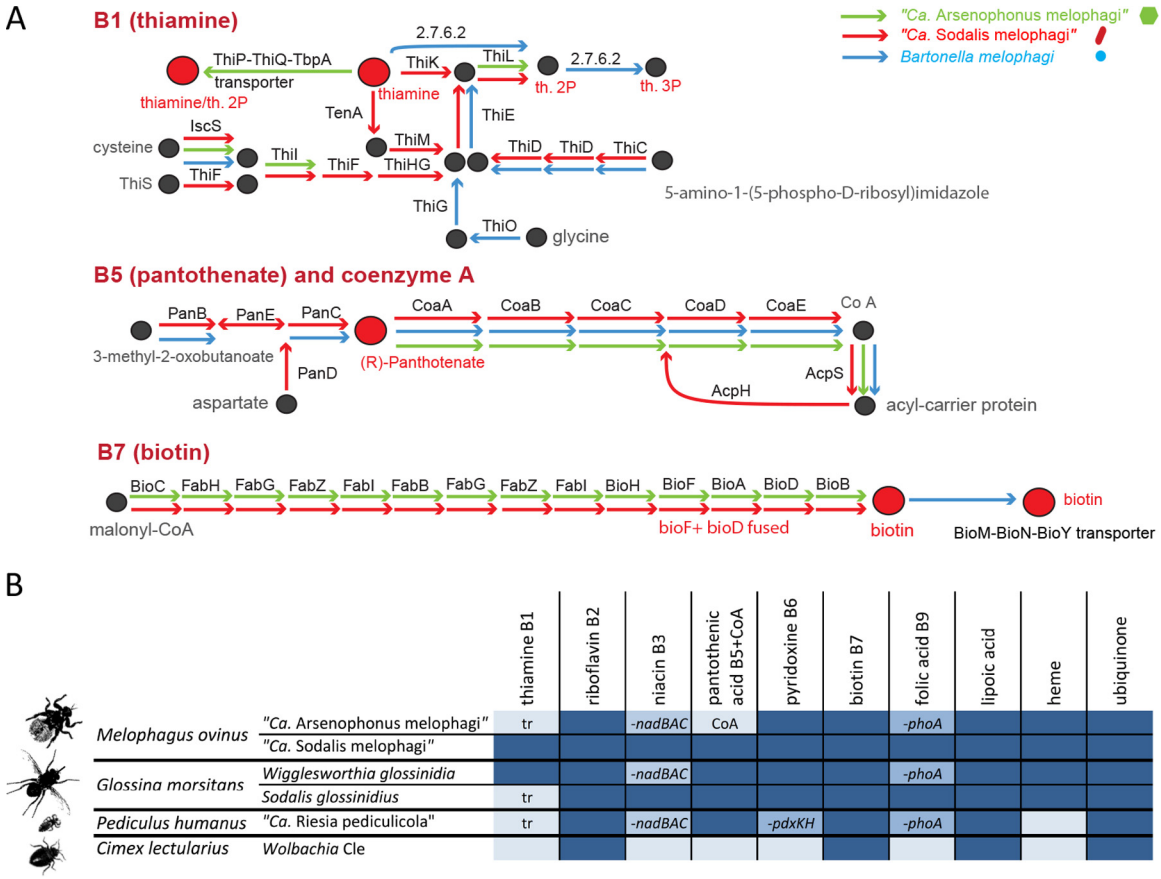

FIG 5 Selected B-vitamin pathways based on draft genomes of *Arsenophonus*, *Sodalis*, and *Bartonella*. (A) The arrows indicate presence of particular genes for different bacteria in *Melophagus ovinus* according to the color legend. (B) The presence of a complete pathway is indicated with the dark blue. Lighter shades are for pathways compromised by missing (−) genes named in the figure. The lightest color indicates corrupted pathways with more than three missing genes. tr, presence of the thiamine ABC transporter.

separate subpopulations in host adult females, one within the bacteriome and the other in the lumen of the milk glands (46–48). A similar arrangement was also reported for the *Arsenophonus*-related symbionts "*Candidatus* Aschnera chinzeii" in the bat fly *Penicillidia jenynsii* (44). It indicates that in these insects the vertical transmission takes place via the accessory organs rather than oocytes, and this mechanism has originated in or been preserved for different groups of bacterial symbionts.

The same transmission path to the developing embryo is also used by the facultative symbionts of the genus *Sodalis*. However, according to our microscope observations, its transmission is not synchronized with that of "*Ca*. Arsenophonus." Rather, "*Ca*. Arsenophonus" seems to colonize the developing embryo prior to "*Ca*. Sodalis." Similarly, the common presence in the host milk gland lumen was only rarely reported for *W. glossinidia* and *S. glossinidius* in Glossinidae (48). This suggests that different mechanisms initiate the vertical transmission of facultative and obligate symbionts in *M. ovinus* or that there is possible competition between them, as proposed previously for the *Glossina* symbionts (48).

The presence of *B. melophagi* along microvilli of the *M. ovinus* midgut indicates possible scavenging of nutrients (e.g., biotin) (from blood digestion and perhaps from the other symbionts (Fig. 5). Although this bacterium, originally cultivated from sheep keds by Noller (49) and described as *Rickettsia melophagi*, was later also cultivated from commercial batches of sheep blood (50), its life cycle is still ambiguous. Halos et al. (51) reported that despite its presence in all *Melophagus* samples collected from a sheep, *B. melophagi* could not be isolated from the sheep itself and explained this by its symbiotic association without transmission to ruminants. Our genomic data do not provide any strong evidence to confirm or rule out this hypothesis. Additional functional data will be needed to infer the exact role of *B. melophagi* in the system.

Glossinidae-Hippoboscidae, a proposed symbiotic model for comparative genomics. While evolving from different ancestors, the two obligate symbionts "*Ca*. Arsenophonus" and *W. glossinidia* have been subject to similar host-mediated influences, reflecting phylogenetic relationships of the two insect groups and their nutritional reliance on vertebrate blood. They thus provide a unique system, potentially useful for distinguishing specific adaptive modifications of the genome from the general degenerative processes. For example, apart from genome reduction, which by itself is hardly surprising as it is a trait generally shared by the obligate symbionts, the two symbionts display high similarity in the pattern of the present/absent orthologous genes (Fig. 4). However, despite this overall similarity, the degree of genomic changes differs across the COG categories (see Tables S2 and S3 in the supplemental material), indicating a possible role of adaptive pro-

cesses. Two patterns may be of particular interest. The first of them is exemplified by the COG category coenzyme transport and metabolism, for which the two hippoboscoid symbionts possess the highest number of genes among all included obligate symbionts. It is interesting that the high gene content of this category is also found in otherwise strongly reduced genomes of "*Ca*. Riesia," a symbiont from sucking lice, and mutualistic *Wolbachia* from bed bugs. This makes the category a promising candidate for the traits possibly important in blood-feeding hosts. A second important pattern is provided by categories where the two symbionts differ considerably from each other in the numbers of present genes. These categories are useful in discriminating the traits that are not associated with the common lifestyle. From this perspective, the difference in the COG category motility is of particular interest. It has been speculated previously that the flagellar system present in the *Wigglesworthia* genome is necessary for transmission of the symbiont via the host milk glands or to enable entry into the larval or pupal bacteriocytes (52). However, in "*Ca*. Arsenophonus," which uses the same transmission route (Fig. 3a and c; see also Fig. S2 in the supplemental material), this COG category is entirely absent. While this finding does not exclude use of the flagella for the transmission in *Wigglesworthia*, it at least indicates that this motility apparatus is not necessary for such a transmission mode. Another striking example for the "*Ca*. Arsenophonus"-*W. glossinidia* comparison is the category energy production and conversion, where the difference of 18 genes is largely due to the presence of an Rnf/Nqr complex and ferredoxin in the "*Ca*. Arsenophonus" genome and its absence in *W. glossinidia*. Similar differences between the two symbionts are also found in some other COG categories, e.g., defense mechanisms, signal transduction mechanisms, and replication, recombination, and repair (Tables S2 and S3 in the supplemental material). The number of genes in the last category is much higher in the "*Ca*. Arsenophonus" genome (as well as the set of genes with unknown functions). This is in good agreement with the perception of "*Ca*. Arsenophonus" as a less degenerate symbiont.

To interpret these observations on a rigorous comparative basis, it will be important to distinguish the variability due to different physiological demands (i.e., adaptive processes) from that reflecting evolutionary age of a given symbiont and/or stochastic processes. This can only been achieved in a broader phylogenetic framework of the hippoboscid-*Arsenophonus* associations using large data sets and proper statistical analysis. The polyphyly of the Hippoboscidae-*Arsenophonus* symbiosis, revealed by 16S rRNA gene- and MLST-based phylogenies (21, 53), suggests its multiple independent origins. This could explain why the genome of "*Ca*. Arsenophonus" is slightly larger than that of *W. glossinidia* if "*Ca*. Arsenophonus" is considered a younger symbiont which originated more recently from a free-living bacterium (possibly replacing other, more ancient symbiont of hippoboscids). In theory the higher number of "*Ca*. Arsenophonus" genes could also be due, at least partially, to gene acquisitions via horizontal gene transmission and/or gene duplications. A simple check showed that a majority of the 725 "*Ca*. Arsenophonus" protein-coding genes returned the best or second-best BLAST hits corresponding to the genus *Arsenophonus* (664) or the closely related genera *Providencia*, *Proteus*, *Photorhabdus*, and *Xenorhabdus* (16). Eighteen hypothetical proteins did not return any hit; for 27 of the remaining genes phylogenetic affinity is less clear and will require more detailed phylogenetic analysis.

Implications for B-vitamin and cofactor biosynthesis in blood-sucking insects. Biosynthesis of B vitamins is traditionally considered a core aspect of the associations between bacteria and blood-sucking insects (13, 43, 54, 55). In our comparison, the arrangement of thiamine biosynthesis and transport presents one of the unexpected metabolic differences between the tsetse fly and louse fly symbiotic systems. In tsetse flies, the genome of *W. glossinidia* codes for the thiamine biosynthetic pathway, and *S. glossinidius* possesses only its ABC transporter (52, 56, 57). In *M. ovinus*, it is completely reversed. Our current data are not sufficient to infer the exact role of thiamine in the system; rather they allow for several not mutually exclusive hypotheses: (i) thiamine is provided to *M. ovinus* and/or "*Ca*. Arsenophonus"by "*Ca*. Sodalis" and/or *B. melophagi*; (ii) "*Ca*. Sodalis" and *B. melophagi* use thiamine solely for their own metabolism, but it is also available from blood *for* "*Ca*. Arsenophonus"; (iii) thiamine supplemented by the symbionts is beneficial but not necessary for host survival (during the whole life cycle or only in certain life stages). Similar hypothetical scenarios may apply to pantothenate biosynthesis: either it is provided by "*Ca*. Sodalis," or it is available from other sources, i.e., sheep blood. This arrangement of the B-vitamin biosynthesis genes and the implied hypotheses indicate that the nutritional interactions between bacteria and blood-feeding insects may be more complex (less determined) than usually supposed. From this perspective, it is interesting that the only other obligate symbiont known to code for thiamine transporter is another member of the *Arsenophonus* clade, "*Ca*. Riesia pediculicola" (58). Since "*Ca*. Riesia" is an exclusive symbiont in human lice, thiamine is likely available from blood and needed by the symbionts for their own metabolism. This uncertainty further highlights our limited knowledge on symbiotic systems of blood feeders. Unlike the symbiotic associations in sap-sucking insects (where symbionts provide their hosts with well-known sets of essential amino acids [59]), it is still quite unclear which B vitamins are provided to blood-sucking hosts and which are needed only for symbionts, with the exception of pyridoxine and thiamine provisioning in the tsetse fly model (11, 14, 15).

Diversity of microorganisms associated with Hippoboscoidea. The system presented here, i.e., the bacteria regularly associated with *M. ovinus*, is an example of the diverse symbiotic associations known from other hippoboscoids. As noted above, the presence of *Arsenophonus* bacteria in different Hippoboscidae members is likely due to multiple origins of the symbiosis or perhaps to recurrent replacements of the obligate symbionts by facultative *Arsenophonus* bacteria (21). The affinity to *Arsenophonus* symbionts is also well demonstrated in a related hippoboscoid group, the bat flies of the family Nycteribiidae (44, 60). Similarly, two other bacterial genera reported here from *M. ovinus*, namely, *Bartonella* and *Wolbachia*, have also been previously identified in other Hippoboscoidea (61–65). *Melophagus ovinus* is, however, the only host known so far with fixed infection of *Bartonella* (51; also this study). Considering this diversity and the abundance of the hippoboscoid-associated symbionts, together with the results presented here for a single hippoboscoid species, this group seems to be a valuable model for further comparative analyses of the insect-bacterium evolution.

ACKNOWLEDGMENTS

This work was supported by the Grant Agency of the Czech Republic (grant 13-01878S to V.H.) and the project Postdoc USB (registration no.

CZ.1.07/2.3.00/30.0006) realized through the EU Education for Competitiveness Operational Programme.

We certify that we have no involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in the manuscript.

REFERENCES

1. **Baumann P.** 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol **59:**155–189. http://dx.doi.org/10.1146/annurev.micro.59.030804.121041.
2. **McCutcheon JP, Moran NA.** 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol **10:**13–26. http://dx.doi.org/10.1038/nrmicro2670.
3. **Moran NA, McCutcheon JP, Nakabachi A.** 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet **42:**165–190. http://dx.doi.org/10.1146/annurev.genet.41.110306.130119.
4. **Petersen FT, Meier R, Kutty SN, Wiegmann BM.** 2007. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. Mol Phylogenet Evol **45:**111–122. http://dx.doi.org/10.1016/j.ympev.2007.04.023.
5. **Meier R, Kotrba M, Ferrar P.** 1999. Ovoviviparity and viviparity in Diptera. Biol Rev **74:**199–258. http://dx.doi.org/10.1017/S0006323199005320.
6. **Lenoble BJ, Denlinger DL.** 1982. The milk gland of the sheep ked, *Melophagus ovinus*: a comparison with *Glossina*. J Insect Physiol **28:**165–172. http://dx.doi.org/10.1016/0022-1910(82)90124-X.
7. **Aksoy S.** 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of mycetocyte associated, primary endosymbionts of tsetse flies. Int J Syst Bacteriol **45:**848–851. http://dx.doi.org/10.1099/00207713-45-4-848.
8. **Cheng Q, Aksoy S.** 1999. Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. Insect Mol Biol **8:**125–132. http://dx.doi.org/10.1046/j.1365-2583.1999.810125.x.
9. **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. http://dx.doi.org/10.1099/00207713-49-1-267.
10. **Aksoy S, Chen X, Hypša V.** 1997. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). Insect Mol Biol **6:**183–190. http://dx.doi.org/10.1111/j.1365-2583.1997.tb00086.x.
11. **Snyder AK, Deberry JW, Runyen-Janecky L, Rio RV.** 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc Biol Sci **277:**2389–2397. http://dx.doi.org/10.1098/rspb.2010.0364.
12. **Nogge G.** 1976. Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. Experientia **32:**995–996. http://dx.doi.org/10.1007/BF01933932.
13. **Nogge G.** 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82:**101–104.
14. **Snyder AK, McLain C, Rio RV.** 2012. The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density via exogenous nutrient provisioning. Appl Environ Microbiol **78:**7792–7797. http://dx.doi.org/10.1128/AEM.02052-12.
15. **Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S.** 2014. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. Appl Environ Microbiol **80:**5844–5853. http://dx.doi.org/10.1128/AEM.01150-14.
16. **Doudoumis V, Alam U, Aksoy E, Abd-Alla AMM, Tsiamis G, Brelsfoard C, Aksoy S, Bourtzis K.** 2013. Tsetse-*Wolbachia* symbiosis: comes of age and has great potential for pest and disease control. J Invertebr Pathol **112:**S94–S103. http://dx.doi.org/10.1016/j.jip.2012.05.010.
17. **Lindh J, Lehane M.** 2011. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie Van Leeuwenhoek **99:**711–720. http://dx.doi.org/10.1007/s10482-010-9546-x.
18. **Aksoy E, Telleria EL, Echodu R, Wu Y, Okedi LM, Weiss BL, Aksoy S, Caccone A.** 2014. Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. Appl Environ Microbiol **80:**4301–4312. http://dx.doi.org/10.1128/AEM.00079-14.
19. **Chrudimský T, Husník F, Nováková E, Hypsa V.** 2012. *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. PLoS One **7:**e40354. http://dx.doi.org/10.1371/journal.pone.0040354.
20. **Nováková E, Hypša V.** 2007. A new *Sodalis* lineage from blood-sucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiol Lett **269:**131–135. http://dx.doi.org/10.1111/j.1574-6968.2006.00620.x.
21. **Nováková E, Hypsa V, Moran NA.** 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol **9:**143. http://dx.doi.org/10.1186/1471-2180-9-143.
22. **Small RW.** 2005. A review of *Melophagus ovinus* (L.), the sheep ked. Vet Parasitol **130:**141–155. http://dx.doi.org/10.1016/j.vetpar.2005.03.005.
23. **Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA.** 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol **56:**1919–1925.
24. **Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH.** 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol **72:**7098–7110. http://dx.doi.org/10.1128/AEM.00731-06.
25. **O'Neill SL, Giordano R, Colbert AME, Karr TL, Robertson HM.** 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci U S A **89:**2699–2702. http://dx.doi.org/10.1073/pnas.89.7.2699.
26. **Norman AF, Regnery R, Jameson P, Greene C, Krause DC.** 1995. Differentiation of *Bartonella*-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. J Clin Microbiol **33:**1797–1803.
27. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
28. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
29. **Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC.** 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res **40:**D115–D122. http://dx.doi.org/10.1093/nar/gkr1044.
30. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
31. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
32. **Lerat E, Daubin V, Moran NA.** 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the γ-*Proteobacteria*. PLoS Biol **1:**e19.
33. **Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH.** 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res **30:**281–283. http://dx.doi.org/10.1093/nar/30.1.281.
34. **Katoh K, Standley DM.** 2014. MAFFT: iterative refinement and additional methods. Methods Mol Biol **1079:**131–146. http://dx.doi.org/10.1007/978-1-62703-646-7_8.
35. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221–224. http://dx.doi.org/10.1093/molbev/msp259.
36. **Castresana J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17:**540–552. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334.
37. **Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods **9:**772–772. http://dx.doi.org/10.1038/nmeth.2109.
38. **Galtier N, Gouy M.** 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol **15:**871–879. http://dx.doi.org/10.1093/oxfordjournals.molbev.a025991.
39. **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic

inference under mixed models. Bioinformatics **19:**1572–1574. http://dx.doi.org/10.1093/bioinformatics/btg180.
40. **Dale JM, Popescu L, Karp PD.** 2010. Machine learning methods for metabolic pathway prediction. BMC Bioinformatics **11:**15. http://dx.doi.org/10.1186/1471-2105-11-15.
41. **Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M.** 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res **40:**D109–D114. http://dx.doi.org/10.1093/nar/gkr988.
42. **Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroeder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD.** 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res **41:**D605–D612. http://dx.doi.org/10.1093/nar/gks1027.
43. **Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T.** 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A **107:**769–774. http://dx.doi.org/10.1073/pnas.0911476107.
44. **Hosokawa T, Nikoh N, Koga R, Sato M, Tanahashi M, Meng XY, Fukatsu T.** 2012. Reductive genome evolution, host-symbiont cospeciation and uterine transmission of endosymbiotic bacteria in bat flies. ISME J **6:**577–587. http://dx.doi.org/10.1038/ismej.2011.125.
45. **Ries E.** 1931. Die Symbiose der Läuse und Federlinge. Zeitschrift für Morphologie und Ökologie der Tiere **20:**233–367.
46. **Aksoy S, Pourhosseeini AA, Chow A.** 1995. Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to *Enterobacteriaceae*. Insect Mol Biol **4:**5–22.
47. **Attardo GM, Lohs C, Heddi A, Alam UH, Yildirim S, Aksoy S.** 2008. Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity. J Insect Physiol **54:**1236–1242. http://dx.doi.org/10.1016/j.jinsphys.2008.06.008.
48. **Balmand S, Lohs C, Aksoy S, Heddi A.** 2013. Tissue distribution and transmission routes for the tsetse fly endosymbionts. J Invertebr Pathol **112:**S116–S122. http://dx.doi.org/10.1016/j.jip.2012.04.002.
49. **Noller W.** 1916. Beitrag zur Flecktyphusubertragung durch Lause. Klin Wochenschr **28:**778.
50. **Bemis DA, Kania SA.** 2007. Isolation of *Bartonella* sp from sheep blood. Emerg Infect Dis **13:**1565–1567. http://dx.doi.org/10.3201/eid1310.070570.
51. **Halos L, Jamal T, Maillard R, Girard B, Guillot J, Chomel B, Vayssier-Taussat M, Boulouis HJ.** 2004. Role of Hippoboscidae flies as potential vectors of *Bartonella* spp. infecting wild and domestic ruminants. Appl Environ Microbiol **70:**6302–6305. http://dx.doi.org/10.1128/AEM.70.10.6302-6305.2004.
52. **Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S.** 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet **32:**402–407. http://dx.doi.org/10.1038/ng986.
53. **Duron O, Schneppat UE, Berthomieu A, Goodman SM, Droz B, Paupy C, Obame Nkoghe J, Rahola N, Tortosa P.** 2014. Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. Mol Ecol **23:**2105–2117. http://dx.doi.org/10.1111/mec.12704.
54. **Puchta O.** 1955. Experimentelle Untersuchungen uber die Bedeutung der symbiose der Kleiderlaus Pediculus vestimenti Burm. Z Parastenkd **17:**1–40. (In German.)
55. **Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T.** 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc Natl Acad Sci U S A **111:**10257–10262. http://dx.doi.org/10.1073/pnas.1409284111.
56. **Belda E, Moya A, Bentley S, Silva FJ.** 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics **11:**449. http://dx.doi.org/10.1186/1471-2164-11-449.
57. **Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, Aksoy S.** 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.
58. **Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li H-M, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, et al.** 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A **107:**12168–12173. http://dx.doi.org/10.1073/pnas.1003379107.
59. **Hansen AK, Moran NA.** 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. Mol Ecol **23:**1473–1496. http://dx.doi.org/10.1111/mec.12421.
60. **Morse SF, Bush SE, Patterson BD, Dick CW, Gruwell ME, Dittmar K.** 2013. Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). Appl Environ Microbiol **79:**2952–2961. http://dx.doi.org/10.1128/AEM.03814-12.
61. **Morse SF, Olival KJ, Kosoy M, Billeter S, Patterson BD, Dick CW, Dittmar K.** 2012. Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). Infect Genet Evol **12:**1717–1723. http://dx.doi.org/10.1016/j.meegid.2012.06.009.
62. **Billeter SA, Hayman DTS, Peel AJ, Baker K, Wood JLN, Cunningham A, Suu-Ire R, Dittmar K, Kosoy MY.** 2012. *Bartonella* species in bat flies (Diptera: Nycteribiidae) from western Africa. Parasitology **139:**324–329. http://dx.doi.org/10.1017/S0031182011002113.
63. **Reeves WK, Loftis AD, Gore JA, Dasch GA.** 2005. Molecular evidence for novel *Bartonella* species in *Trichobius major* (Diptera: Streblidae) and *Cimex adjunctus* (Hemiptera: Cimicidae) from two southeastern bat caves, USA. J Vector Ecol **30:**339–341.
64. **Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, O'Neill SL, Aksoy S.** 2000. Tissue distribution and prevalence of *Wolbachia infections in tsetse flies*, *Glossina* spp. Med Vet Entomol **14:**44–50. http://dx.doi.org/10.1046/j.1365-2915.2000.00202.x.
65. **Doudoumis V, Tsiamis G, Wamwiri F, Brelsfoard C, Alam U, Aksoy E, Dalaperas S, Abd-Alla A, Ouma J, Takac P, Aksoy S, Bourtzis K.** 2012. Detection and characterization of *Wolbachia* infections in laboratory and natural populations of different species of tsetse flies (genus *Glossina*). BMC Microbiol **12:**S3. http://dx.doi.org/10.1186/1471-2180-12-S1-S3.
66. Reference deleted.
67. **Dale C, Young SA, Haydon DT, Welburn SC.** 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci U S A **98:**1883–1888. http://dx.doi.org/10.1073/pnas.98.4.1883.
68. **Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S.** 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. http://dx.doi.org/10.1101/gr.4106106.