Insect—Symbiont Gene Expression in the Midgut Bacteriocytes of a Blood-Sucking Parasite

Filip Husnik [1,2,*,†], Vaclav Hypsa [1,2], and Alistair Darby[3]

[1]Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, České Budějovice, Czech Republic

[2]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

[3]Institute of Integrative Biology, University of Liverpool, United Kingdom

[†]Present address: Okinawa Institute of Science and Technology, Okinawa, Japan

*Corresponding author: E-mail: filip.husnik@gmail.com.

Abstract

Animals interact with a diverse array of both beneficial and detrimental microorganisms. In insects, these symbioses in many cases allow feeding on nutritionally unbalanced diets. It is, however, still not clear how are obligate symbioses maintained at the cellular level for up to several hundred million years. Exact mechanisms driving host–symbiont interactions are only understood for a handful of model species and data on blood-feeding hosts with intracellular bacteria are particularly scarce. Here, we analyzed interactions between an obligately blood-sucking parasite of sheep, the louse fly *Melophagus ovinus*, and its obligate endosymbiont, *Arsenophonus melophagi*. We assembled a reference transcriptome for the insect host and used dual RNA-Seq with five biological replicates to compare expression in the midgut cells specialized for housing symbiotic bacteria (bacteriocytes) to the rest of the gut (foregut–hindgut). We found strong evidence for the importance of zinc in the system likely caused by symbionts using zinc-dependent proteases when acquiring amino acids, and for different immunity mechanisms controlling the symbionts than in closely related tsetse flies. Our results show that cellular and nutritional interactions between this blood-sucking insect and its symbionts are less intimate than what was previously found in most plant-sap sucking insects. This finding is likely interconnected to several features observed in symbionts in blood-sucking arthropods, particularly their midgut intracellular localization, intracytoplasmic presence, less severe genome reduction, and relatively recent associations caused by frequent evolutionary losses and replacements.

Key words: RNA-Seq, B-vitamins, parasites, symbiotic bacteria, interactions, zinc, immunity.

Introduction

Nutritional supplementation from symbiotic bacteria allowed several insect groups to specialize on nutritionally unbalanced diets. Endosymbiotic bacteria provide essential amino-acids and some B-vitamins to plant-sap sucking hosts, B-vitamins to blood-sucking hosts, and assist wood-feeding insects with nitrogen recycling (Moran et al. 2008; McCutcheon and Moran 2012; Douglas 2016). What is still unclear is if there are any general insect-wide mechanisms driving these host–symbiont interactions. For example, how often do independent solutions of exactly the same functional problem emerge in even closely related species? To what extent do these interactions depend on the host biology (e.g., plant-sap vs. blood-sucking insects) rather than on the host phylogeny (e.g., blood-sucking Hemiptera vs. Diptera) or the symbiont phylogeny (e.g., Gammaproteobacteria vs. Bacteroidetes)? Because the small symbiont genomes are usually subsets of well-known bacterial genomes such as *Escherichia coli*, it is feasible to roughly predict their metabolism either by simple mapping of present genes on metabolic pathways (Hansen and Moran 2014; Husnik and McCutcheon 2016) or by systems biology approaches such as Flux Balance Analysis (Thomas et al. 2009; MacDonald et al. 2011; Belda et al. 2012; Gonzalez-Domenech et al. 2012). What is not easily

feasible, though, is to fully encompass and model the interactions with the host, especially for nonmodel species for which host genome, transcriptome or proteome data are not available.

Our Understanding of Nutritional Insect-Symbiont Provisioning Comes Mostly from Plant-Sap Sucking Insects

Majority of data concerning the host role in arthropod-bacteria symbiosis is undoubtedly available for pea aphids (Nakabachi et al. 2005; Gerardo and Wilson 2011; Hansen and Moran 2011; Poliakov et al. 2011). Hansen and Moran (2011) and Poliakov et al. (2011) untangled the intimate symbiotic interface in the pea aphid-*Buchnera* system, and confirmed the previously suggested (Nakabachi et al. 2005) host–symbiont cooperation in the production of essential amino acids (EAAs) and incorporation of ammonium nitrogen into glutamate. Additional data (Macdonald et al. 2012) show that waste ammonia is recycled predominantly by the host cells and that aphid aminotransferases (ornithine AT: EC 2.6.1.13, branched-chain AT: EC 2.6.1.42, and aspartate AT: EC 2.6.1.1) incorporate ammonia-derived nitrogen into carbon skeletons synthesized by *Buchnera* to generate EAAs. The highly similar picture observed in citrus mealybugs (Husnik et al. 2013), petiole gall psyllids (Sloan et al. 2014) or whiteflies (Luan et al. 2015), and identical enzymatic gaps in other endosymbiont genomes from hemipterans (Hansen and Moran 2014), imply that many (but perhaps not all; Van Leuven et al. 2014) insect hosts carry out these last steps to gain control of production of the final products.

Unlike animals, plants can synthesize B-vitamins (Roje 2007), but whether B-vitamins are acquired by insects from the phloem/xylem sap of their host plants and provided to endosymbionts remains poorly understood. Endosymbiont genomes from plant-sap feeding insects retain several genes/pathways for biosynthesis of B-vitamins, for example, biotin, riboflavin, and folate (Hansen and Moran 2014; Moran and Bennett 2014). Which B-vitamins are only used by symbionts and which are in addition also provided to their hosts is, however, unclear. The only piece of experimental evidence implies that young symbiotic aphids are provided with riboflavin by their *Buchnera* endosymbionts (Nakabachi and Ishikawa 1999) and riboflavin provision is also implicated in aphid co-obligate symbioses (Manzano-Marín et al. 2016).

Host–Symbiont Cooperation Depends on Transport of Compounds between the Bacteriocytes and the Symbiont Cells

Symbiotic bacteria of plant-sap sucking insects retain only a few general transporters, some of which very likely lost their substrate specificity (Charles et al. 2011). On the other hand, the host transporters can be involved in symbiont maintenance. For example, amino acid transporters of sap-feeding insects were extensively duplicated and specialized for bacteriocyte transfer (Duncan et al. 2014) and symbiont control (Price et al. 2014; Lu et al. 2016). No evidence of massive transfer of proteins among the symbiotic partners was so far confirmed, although one host protein was reported to be targeted to *Buchnera* cells in aphids (Nakabachi et al. 2014). However, such protein transfer is very likely needed in other hosts. For example, a recent rigorous analysis of host expression in two bacteriome types in a leafhopper host implies that nucleus-encoded genes usually supporting mitochondria also support bacterial endosymbionts (Mao et al. 2018).

Nutritional Interactions between Blood-Sucking Insects and Their Symbiotic Bacteria Are Understood Only for a Few Hosts

Based on genomic data, different bacterial symbionts of blood-feeding insects can synthesize biotin, thiamine, riboflavin and FAD, panthotenate and coenzyme A, folate, pyridoxine, ubiquinol, nicotinamide, lipoic acid, and protoheme (Kirkness et al. 2010; Rio et al. 2012; Nikoh et al. 2014; Nováková et al. 2015; Boyd et al. 2016; Říhová et al. 2017). Controversy arises when discussing which particular cofactors are provided in particular host lineages. Interestingly, there are obligately blood-feeding arthropods (e.g., ticks or kissing bugs) that do not house stable intracellular microbes. These athropods either efficiently extract rare nutrients from their blood diet or rely on extracellular gut bacteria acquired from the environment, for example, by coprophagy as in kissing bugs (Eichler and Schaub 2002). Different blood-feeding lineages thus likely rely on symbionts for different subsets of these cofactors, perhaps due to differences in their blood-feeding strategies, association with the host, host species, enzymatic dependence (e.g., using alternative enzymes not depending on a particular cofactor), and evolutionary history. Some of the cofactors produced by symbionts are likely only used by symbiont-encoded enzymes rather than being provided to the insect host. Other cofactors such as thiamine in human lice and louse flies are acquired from the blood diet by the host and made available to the symbionts which extract it by their thiamine ABC transporters (reviewed in Husnik 2018).

In comparison to highly mosaic pathways such as biosynthesis of branched-chain amino acids in plant-feeding insects where host enzymes are needed to synthesize the final product, bacteria in blood-feeding insects seem to be functioning more as independent units. The only RNA-Seq analysis from blood-feeding insects with intracellular symbionts was carried out in tsetse flies. The authors show that in terms of nutritional cooperation, only a few host genes seem to maintain the symbiosis, particularly a multi-vitamin transporter is upregulated to shuttle B-vitamins from midgut bacteriocytes to hemolymph or other tissues (Bing et al. 2017).

Insect Immune Response Often Distinguishes Obligate Mutualists from Facultative Symbionts and Pathogens

Several ancient and intracellular obligate symbionts of insects have partially or completely lost bacterial cell envelope structures recognized by the insect immune system—peptidoglycan and lipopolysaccharides (McCutcheon and Moran 2012). In the latter case, they are often engulfed by a host-derived symbiosomal membrane (McCutcheon and Moran 2012), so there is nothing on their cell envelopes recognized as of bacterial origin by the host peptidoglycan-recognition proteins (PGRPs) or Gram-negative binding proteins (GNBPs). Interestingly, if there are some structures of bacterial origin still present, the hosts were either shown to jettison PGRPs, genes from the immunodeficiency signaling pathway (IMD), and many antimicrobial peptides (Gerardo et al. 2010; Kirkness et al. 2010) or modify them for symbiont defense. Blood-feeding insects that still keep PGRPs such as tsetse flies use amidase activity of one of PGRPs for peptidoglycan recycling in bacteriocytes (and milk glands of tsetse flies) and this activity shields symbionts from recognition by other PGRPs and expression of lineage-specific antimicrobial peptides mediated by the IMD (Anselme et al. 2006; Anselme et al. 2008; Wang et al. 2009; Weiss et al. 2011; Wang and Aksoy 2012; Ratzka et al. 2013; Bing et al. 2017). In *Sitophilus* weevils, a recently "domesticated" *Sodalis pierantonius* symbiont (Oakeson et al. 2014) was shown to be kept under control by a single antimicrobial peptide (Coleoptericin A). After silencing its expression by RNA interference, the symbionts escaped from the bacteriocytes and spread into host tissues (Login et al. 2011). The functional role of the symbionts is to synthesize a single non-EAA, tyrosin, and when this benefit is no longer needed (for cuticle hardness), symbiont numbers are reduced by the host (Vigneron et al. 2014). *Sodalis*-allied bacteria are very common symbionts of blood-sucking insects, so it is likely that similar symbiont control solutions had to also emerge in blood-feeding insects.

Melophagus ovinus is a sheep ectoparasite highly adapted for blood-feeding. Both females and males are wingless, permanently associated with their sheep host, and strictly blood-feeding. Similarly to other Hippoboscoidea such as tsetse flies, *M. ovinus* is fully dependent on intracellular symbiotic bacteria for its survival/reproduction, houses the bacteria in midgut bacteriocytes, and the symbionts are transmitted from the mother to its intrauterine larvae through the secretions of specialized "milk" glands (adenotrophic viviparity). One obligate symbiotic bacterium, *Arsenophonus melophagi*, is always present in all *Melophagus* individuals whereas some host populations can be also infected by facultative bacteria *Sodalis melophagi*, *Bartonella melophagi*, and *Wolbachia* sp.; and the facultative eukaryote *Trypanosoma melophagium* (Small 2005; Chrudimský et al. 2012; Nováková et al. 2015; Nováková et al. 2016). Here, we use comprehensive dual host–symbiont RNA-Seq data for *M. ovinus*, and its obligate

symbiont, *A. melophagi*, to understand metabolic interdependence and host–symbiont interactions.

Materials and Methods

RNA Extraction and Sequencing

Melophagus ovinus parasites were sampled from the same population as used for our previous studies (Chrudimský et al. 2012; Nováková et al. 2015). Insects were immediately dissected in RNAlater (Qiagen) to stabilize expression profiles and kept deeply frozen until RNA extractions. Five biological replicates of bacteriomes and five replicates of the remaining portion of gut from identical samples were prepared for RNA-Seq. Total RNA was extracted from pools of five individual bacteriomes and guts for seven replicates using RNeasy Mini Kit (Qiagen). Five of the replicates for both tissues were selected for sequencing based on RNA quality and Bioanalyzer chip results (Agilent). All samples were DNased by RNase-free DNase I (Ambion) and ribosomal RNA was depleted by a Terminator-5-Phosphate-Dependent Exonuclease treatment (Epicentre). Ten RNA-Seq libraries were prepared from the enriched RNA samples using the ScriptSeq strand-specific protocol (Epicentre). Paired-end sequencing (2 × 100 bp) of the ten RNA-Seq libraries was multiplexed on one lane of the Illumina HiSeq 2000 platform at the Centre for Genomic Research, University of Liverpool. Total RNA was also extracted from one whole female and one whole male *M. ovinus* to improve de novo assembly of the reference transcriptome. The identical procedure as described above was carried out and the two libraries were multiplexed on one lane of the HiSeq 2000. Raw Fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt v1.1 (https://cutadapt.readthedocs.io/, last accessed February 25, 2020). Option -O 3 was used, so that the 3' end of any read which matched the adapter sequence for 3 bp or more was trimmed. The reads were further quality-trimmed by Sickle v1.200 (https://github.com/najoshi/sickle, last accessed February 25, 2020) with a minimum window quality score of 20. Reads <10 bp after trimming and singlet reads were removed. The quality-trimming resulted in 371,432,771 read pairs. PhyloFlash v3.1 (Gruber-Vodicka et al. 2019) was used for rRNA depletion efficiency and contamination assessment using 18S rRNA and 16S rRNA gene databases.

Arsenophonus melophagi Gene Expression and Analyses

Bacterial expression was analyzed by mapping reads from individual libraries to the *A. melophagi* assembly with Bowtie2 (Langmead and Salzberg 2012). Bam files were imported into BamView integrated in the Artemis browser v15.1.1 (Carver et al. 2012). The *A. melophagi* annotation (available from Nováková et al. 2015; http://users.prf.jcu.cz/novake01/, last accessed February 25, 2020) was subsequently improved according to the expression data, that is, mainly pseudogene

remnants (not-expressed short hypothetical proteins) were re-annotated. These pseudogene re-annotation results were also supported by our custom pipeline (https://github.com/filip-husnik/pseudo-finder, last accessed February 25, 2020). Identification of transcript boundaries, quantification of transcript abundance (FPKM values), and prediction of operon structure was carried out in Rockhopper v2.0.3 using default parameters (McClure et al. 2013).

De Novo Metatranscriptome Assembly and Differential Expression Analyses

De novo metatranscriptome assembly was carried out by the Trinity assembler v2.4.0 (Grabherr et al. 2011) from digitally normalized read pairs (targeted maximum coverage set to 50) using strand-specific information. RSEM 1.3.0 (Li and Dewey 2011) was used to count Bowtie2-mapped (nonnormalized) reads. The trimmed mean of M-values normalization (TMM), generation of normalized TPM values (transcripts per million transcripts), and differential expression analyses with five biological replicates were carried out in EdgeR (Robinson et al. 2010) BioConductor package using settings recommended for transcriptome assemblies generated by the Trinity assembler (run_DE_analysis.pl script). Only transcripts with at least 4-fold expression change (P-value cut-off for false discovery rate set to 0.001) were considered to be differentially expressed between the bacteriome and gut tissues.

Reference Transcriptome Filtering for Bacterial and Eukaryotic Contamination

The reference transcriptome of *M. ovinus* was filtered for likely assembly artifacts and lowly supported transcripts so that only transcripts with normalized TPM > 1 in at least one sample or replicate were retained (retained 30.95%, i.e., 51,386 from 166,038 of total transcripts). Taxonomy was assigned to all transcripts by Blobtools (Laetsch and Blaxter 2017) with the "−bestsum" flag based on BlastN searches against the NT database and Diamond BlastX against the Uniprot proteome database. All transcripts assigned either to the superkingdom Bacteria (symbionts and bacterial contamination), the phylum Chordata (sheep and human contamination), or the Kinetoplastida class (*T. melophagium*) were removed from the reference transcriptome. However, we note that some of *T. melophagium* transcripts could be missed by this approach because transcripts with no hits were retained. Single-gene phylogenetic inference would be needed to fully decontaminate *T. melophagium* contamination from the *Melophagus* reference transcriptome because no reference genome exists for neither *M. ovinus* nor *T. melophagium*.

Transcriptome Annotation and Protein Prediction

TransDecoder (https://github.com/TransDecoder/, last accessed February 25, 2020) was used for ORF and protein prediction of the filtered reference proteome. Complete functional annotation was produced by Trinotate and final results were uploaded into a custom MySQL database and analyzed through TrinotateWeb (Grabherr et al. 2011). BUSCO v3 was used to assess the transcriptome completeness against 303 universally conserved eukaryotic proteins (Simão et al. 2015) both before and after contamination removal.

Reconstruction of Selected Pathways, Metabolite Exchange, and Immune System Composition

Digital expression values were overlaid on *A. melophagi* pathway map in the Pathway Tools Software (Karp et al. 2010). Up-regulated host transcripts of interest (e.g., B-vitamin, co-factor, metal, and amino acid metabolism) were analyzed manually for possible interactions with symbionts. Previously published pathways for B-vitamins were updated with this expression information. Expression of genes annotated in the *A. melophagi* genome as transporters was assessed to predict metabolite flux to and from the symbionts. Similarly, annotations of up-regulated host transcripts were screened for transporters to analyze the host role in the maintenance of symbiotic tissue. Transporter candidates were checked by BlastP (e-value $1e^{-6}$) against the NR protein database. EMBL-EBI CoFactor database (Fischer et al. 2010) was used for searches of cofactor enzyme dependence.

The insect immune system was reconstructed by BLAST searches (BlastP, BlastN, and BlastX) of homologs against a custom BLAST database built from our transcriptome assembly and gene candidates were tested against the nonredundant NCBI database. In particular, *Drosophila melanogaster* and *Glossina morsitans* homologs acquired from the Insect Innate Immunity Database (Brucker et al. 2012) and VectorBase (https://www.vectorbase.org/, last accessed February 25, 2020) were used as queries. Pathways involved in control of symbiotic bacteria in gut and bacteriome tissues were compared with the situation in tsetse flies using literature review.

Results

Contamination and rRNA Depletion Efficiency Assessment

Because ribosomal RNA was depleted by a Terminator-5′-Phosphate-Dependent Exonuclease treatment to allow parallel evaluation of expression for both the host and all its microbiome members, we tested the depletion efficiency by mapping trimmed reads against comprehensive 18S rRNA and 16S rRNA [small ribosomal subunit (SSU)] gene databases using PhyloFlash. Approximately 27–30% of all reads per library mapped to these genes. However, even if large subunit rRNA molecules (23S and 28S rRNA) make up more of the data than the small subunit rRNAs, total amount of sequencing data generated (371,432,771 trimmed reads) not only allowed us to assemble a high-quality reference transcriptome for *M. ovinus*, but also to carry out abundance estimation and

A. Ribosomal RNA depletion efficiency and taxonomic summary from SSU reads mapped
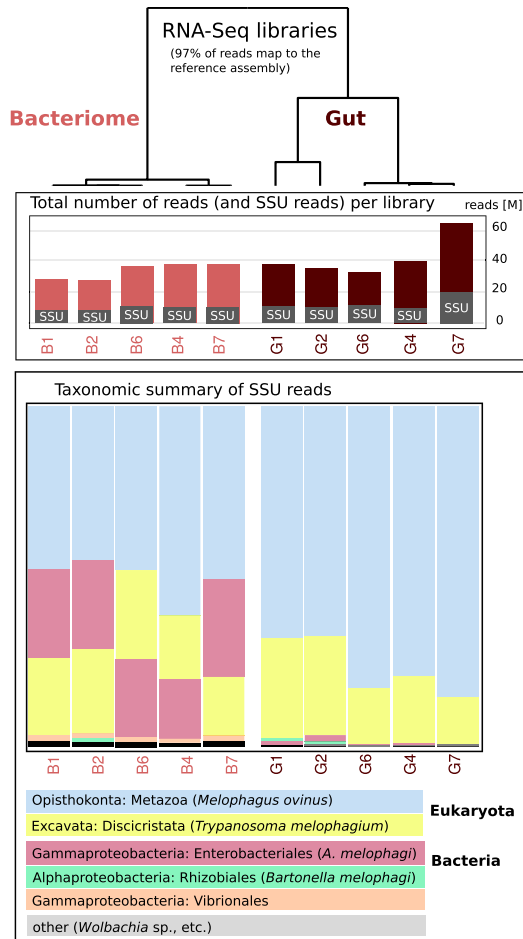
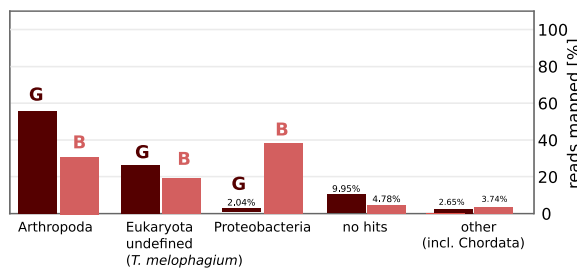## B. Taxonomic summary from the total metatranscriptome

FIG. 1.—Ribosomal RNA depletion efficiency and species composition/contamination assessment. (A) The first barplot shows total number of reads per library and proportion of SSU reads detected in those. The second chart represents only the small ribosomal rRNA reads (both bacterial and eukaryotic) assigned to taxonomic categories and sorted based on proportion from lowest to highest. (B) The barplot shows percentage of reads mapped to taxon-annotated transcripts from the total meta-transcriptome data.

expression analyses for *M. ovinus* and *A. melophagi*. Species composition profiles allowed by SSU mapping revealed very homogenous microbiome composition across all samples (fig. 1A) with majority of data from the host *M. ovinus* (and

low coverage sheep contamination from blood), its obligate bacterial symbionts *A. melophagi*, and a eukaryotic commensal/pathogen *T. melophagium*. The results show that the method used for rRNA depletion had relatively low efficiency, but also clearly confirm that *A. melophagi* is the most dominant symbiont housed in the midgut bacteriocytes. Additional lower abundance taxa of interest include *S. melophagi*, *B. melophagi*, *Wolbachia* sp., and an unidentified ciliate related to sheep rumen ciliates. Species composition evaluation of the assembled metatranscriptome in Blobtools further corroborated these SSU results (fig. 1B). We note that ~97% of all reads map to our metatranscriptome assembly.

Evaluation of Symbiont Expression

FPKM values obtained by mapping raw data on the *A. melophagi* genome are presented in supplementary 1, Supplementary Material online and the improved *A. melophagi* annotation is available through FigShare (10.6084/m9.figshare.6146777). Figure 2 shows the expression values and predicted operon structure of the *A. melophagi* genome overlaid on its linear map with genes color-coded according to their COG (Clusters of Orthologous Groups) functional assignments. Expression values for the 20 most highly expressed *A. melophagi* genes are shown in table 1.

De Novo Transcriptome Assembly and Differential Expression Analyses of Host Data

The transcriptomic data were de novo assembled into a total of 166,038 transcripts (83,574 "genes") with a contig N50 size of 1,456 bp. Total 166,522,714 bases were assembled. BUSCO completeness assessment resulted in 298 complete (202 duplicates due to the presence of *T. melophagium* and sheep transcripts) and 5 fragmented genes out of 303 markers. Transdecoder ORF prediction resulted in 98,820 proteins (31,807 full-length; 35,448 partial; 31,565 internal) providing us with a robust proteome set for analysis of the host role in the symbiotic system. Our strictly decontaminated *M. ovinus* transcriptome (no transcripts with TPM < 1 or taxonomic assignment to Bacteria, Kinetoplastida or Chordata by Blobtools) consists of 43,315 transcripts with 23,780 predicted proteins (9,037 complete). These filtered data had the BUSCO score of 272 complete (80 duplicates), 8 fragmented, and 23 missing markers (out of 303 markers).

Raw counts of mapped reads, normalized digital expression values, and results of differential expression analyses from EdgeR for gut and bacteriome tissues are available in supplementary 2, Supplementary Material online. In total, 249 genes were found significantly up-regulated and 266 genes significantly down-regulated in the bacteriome section of the midgut (supplementary 2, Supplementary Material online). We note that the total number of up-regulated genes detected by EdgeR was 1353 due to *Arsenophonus* transcripts that were not excluded from the total assembly for
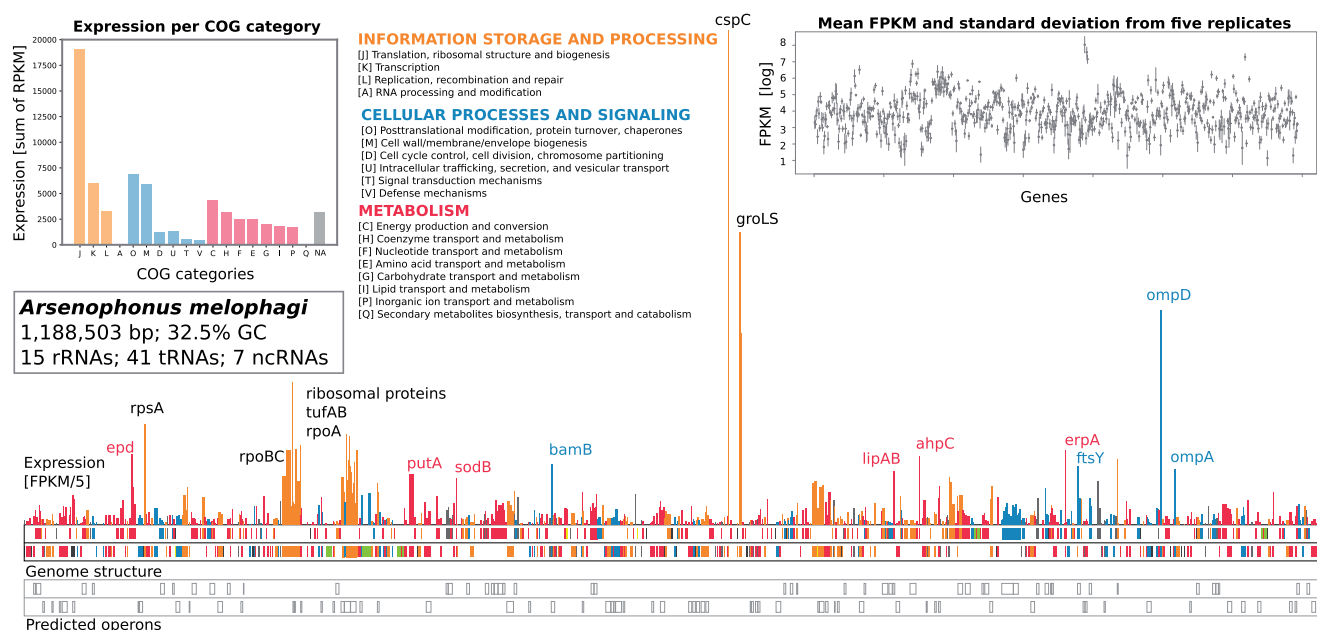
Fig. 2.—Linear genome map of *Arsenophonus melophagi* genome overlaid by expression values of protein-coding genes. Genes are color-coded according to their broad COG functional categories: Information storage and processing in dark orange, cellular processing and signaling in steel blue, and metabolism in crimson. The most highly expressed genes (other than translation-related ribosomal proteins) are highlighted by gene names next to their respective peaks. Ribosomal RNA genes are shown in light green and tRNA/ncRNA genes in yellow. Expression of rRNAs is not shown because it was depleted and is likely biased. Expression of tRNAs and other short ncRNAs is not shown because these RNAs are underrepresented due to RNA isolation and library preparation methods used. Top left inset: Total sum of FPKM expression values for individual categories is included as a bar plot. Genes assigned to two or more categories were not included for simplicity. Top right inset: Mean FPKM values with standard deviation showing consistent expression across five biological replicates of bacteriomes. The FPKM values are plotted on a log scale. Bottom: Genes predicted to be expressed in operons.

the differential expression step. These transcripts (see supplementary 2, Supplementary Material online) are not considered here because *Arsenophonus* is only present in the bacteriome. The most highly up-/down-regulated genes that have functional annotation are shown in table 2.

Interestingly, 11 transcripts significantly down-regulated in the bacteriome tissue come from unclassified RNA viruses (distantly related to the Hubei Diptera virus 14). Unfortunately, these transcripts are short and fragmented, perhaps due to their nucleotide diversity. Although an experimental verification of this finding is needed, it is tempting to speculate that either endosymbionts in the bacteriome tissue (*A. melophagi* and *S. melophagi*) protect from these viruses or that endosymbionts localized mostly outside the bacteriome tissue (*Wolbachia pipientis* and *B. melophagi*) promote the viral infection.

Reconstruction of Selected Pathways Involved in Host–Symbiont–Pathogen Interactions

Cofactor metabolism reconstruction uncovered that *A. melophagi* not only synthesizes several B-vitamins (B2, B3, B6, B7, B9) in bacteriocytes, but that it also acquires thiamine (B1) and pantothenate (B5) from the host cells. Possible exploitation of this resource by the additional members of the

microbiome, that is, *B. melophagi*, *S. melophagi*, *Wolbachia* sp., and *T. melophagium* is also likely (fig. 4), because gene expression data for these facultative microbiome members were of relatively low abundance, we do not draw any strong conclusions from these data here. Host immune system reconstruction and identification of genes differentially expressed in the bacteriome and midgut sections provide evidence for a possible mechanism of how the host keeps its bacterial symbionts under control and how could facultative members of the microbiome escape its recognition (fig. 4, Suppl 4, Supplementary Material online). Gene/transcript information used for the reconstruction of the metabolic pathways and immune system (supplementary 3, Supplementary Material online) can be found in supplementary 1 and 2, Supplementary Material online. Additional basic statistics including the total number of reads (after quality trimming) generated per library and per tissue/sample, and the total number of reads mapped to the reference metatranscriptome and genomes of the four most abundant symbionts are available in supplementary 4, Supplementary Material online.

Transporters Up-Regulated in Bacteriocytes

Two host transporters possibly involved in maintenance of the symbiotic system were found to be significantly up-regulated

Table 1

Thirty Most Highly Expressed *Arsenophonus melophagi* Genes

Gene	Product	COG	FPKM
cspC	Cold shock-like protein	K	3418
groL	60 kDa chaperonin	O	2025
ompD	Outer membrane porin	M	1487
groS	10 kDa chaperonin	O	1322
rplJ	50S ribosomal protein L10	J	986
rpsA	30S ribosomal protein S1	J	693
rpmD	50S ribosomal protein L30	J	628
rplX	50S ribosomal protein L24	J	609
rpsL	30S ribosomal protein S12	J	553
tufA_1	Elongation factor Tu	J	524
rpoB	RNA polymerase subunit β	K	518
tufA_2	Elongation factor Tu	J	512
erpA	Iron-sulfur cluster protein	C	512
epd	D-erythrose-4-P dehydrogenase	G	486
ahpC	Alkyl hydroperoxide reductase subunit C	O	475
rplM	50S ribosomal protein L13	J	453
rpsN	30S ribosomal protein S14	J	447
rplW	50S ribosomal protein L23	J	430
bamB	Outer membrane protein	M	416
rplN	50S ribosomal protein L14	J	408
ftsY	Signal recognition particle receptor	U	405
ompA	Outer membrane protein A	M	386
lipA	Lipoyl synthase	H	371
rplP	50S ribosomal protein L16	J	365
putA	Bifunctional protein PutA	E	352
rpoC	RNA polymerase subunit β'	K	336
rpsM	30S ribosomal protein S13	J	332
rpsU	30S ribosomal protein S21	J	330
rpoA	RNA polymerase subunit α	K	327
sodB	Superoxide dismutase [Fe]	P	325

Note.—COG functional categories: M, cell wall/membrane/envelope biogenesis; O, posttranslational modification, protein turnover, and chaperones; U, intracellular trafficking, secretion, and vesicular transport; J, translation, ribosomal structure and biogenesis; K, transcription; C, energy production and conversion; E, amino acid transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; P, inorganic ion transport and metabolism. FPKM values are averaged from five replicates.

in bacteriocytes (table 2) and likely regulating transfer of zinc and trehalose to the cytoplasm of the bacteriocyte cells housing *A. melophagi.* Other host transporters found to be highly expressed along the entire gut include, for example, a sodium-dependent multivitamin transporter, a mitochondrial sodium-hydrogen exchanger, a sodium-potassium ATPase subunit alpha, a proton-coupled amino acid transporter, and a copper uptake protein.

The expressed transporters of *A. melophagi* are highlighted in figure 3. Notably, thiamine and pantothenate transporters suggest that *A. melophagi* imports these two cofactors into its cells from the host blood diet or other microbiome members.

Additional Transcripts of Bacterial Origin

All transcripts recognized by our cut-off values as of bacterial origin are available in supplementary 2, Supplementary

Material online. Most of these transcripts originated from endosymbionts present in the host species in both the gut and bacteriome tissues. The remaining bacterial transcripts represent low expression contamination by environmental and gut bacteria, false positive eukaryotic transcripts, or possible HGT candidates in the *M. ovinus* or *T. melophagium* genomes (supplementary 2, Supplementary Material online). No additional facultative symbiont species such as *Rickettsia* or *Cardinium* spp. were found to be present in our *M. ovinus* data by any of the methods employed (PhyloFlash, Blobtools, and BlastN/BlastX filtering). We do not discuss these genes of possible bacterial origin further because much more contiguous genome data for both the host and its facultative microbiome members would be needed to reliably distinguish horizontal gene transfer events in the *M. ovinus* genome and microbiome.

Discussion

Arsenophonus Gene Expression and Biosynthesis of B-Vitamins and Cofactors

The expression profile of *A. melophagi* (fig. 2; supplementary 1, Supplementary Material online; table 1) represents a typical example of an obligate symbiont in an intermediate stage of genome reduction. As expected for such a symbiont genome, genes for essential bacterial machinery (replication, transcription, translation) components such as ribosomal proteins, elongation factors, and RNA and DNA polymerases were among the most highly expressed. The most highly expressed gene was the cold shock-like protein CspC (and its antisense ncRNA) with transcription antiterminator activity. Other highly expressed genes were found to be for heat shock proteins (GroEL and GroES; EC 5.6.1.7), outer membrane proteins, and porins allowing various small solutes to cross the outer membrane (OmpD, OmpA, BamABDE), enzymes protecting against reactive oxygen species such as alkyl hydroperoxide reductase (EC 1.11.1.-) and superoxide dismutase (EC 1.15.1.1), and iron-sulfur cluster insertion protein ErpA (fig. 2; supplementary 1, Supplementary Material online). Very similar sets of highly expressed genes were also found in other obligate endosymbionts (Fukatsu and Ishikawa 1993; Aksoy 1995; Baumann et al. 1996; Fan et al. 2013). A pathway that shows unexpectedly high expression is conversion of proline to L-glutamate through the PutA enzyme (EC 1.5.5.2/1.2.1.88) and its subsequent conversion to D-glutamate by the MurI enzyme (EC 5.1.1.3). Considering that proline is almost always the most common amino acid in insect hemolymph (Arrese and Soulages 2010), it is likely that symbionts use proline-derived D-glutamate for their peptidoglycan cell wall. Proline is in insects generally stored as an energy reserve for energy-demanding activities (e.g., flight) or for adaptation to cold temperature, however, *M. ovinus* is flightless and permanently associated with its warm-blooded host. Its proline reserves can be thus potentially used by symbiotic bacteria

Table 2

Significantly (A) Up-Regulated Host Transcripts in the Bacteriome Tissue and (B) Down-Regulated Host Transcripts in the Bacteriome Tissue (Compared with the Rest of the Gut)

Annotation [Transcript ID]	Fold Change		
	logFC	P Value	FDR
(A) Up-regulated transcripts			
FLYWCH-type zinc finger-containing protein 1 [DN33516_c0_g1_i7]	5.792	1.86×10^{-12}	1.32×10^{-10}
FLYWCH-type zinc finger-containing protein 1 [DN33516_c0_g1_i2]	4.708	4.07×10^{-09}	1.90×10^{-07}
FLYWCH-type zinc finger-containing protein 1 [DN33516_c0_g1_i5]	3.941	1.60×10^{-09}	8.01×10^{-08}
Tyrosine-protein phosphatase 10D [DN35320_c1_g6_i1]	5.325	1.12×10^{-16}	1.42×10^{-14}
Paramyosin, short form isoform [DN33516_c0_g1_i2]	4.645	1.30×10^{-10}	7.38×10^{-09}
Zinc transporter foi [DN36258_c3_g8_i2]	4.082	4.90×10^{-10}	2.64×10^{-08}
Zinc transporter foi [DN37931_c0_g1_i2]	3.430	6.74×10^{-10}	3.58×10^{-08}
Glutathione-specific gamma-glutamylcyclotransferase 1 [DN33160_c6_g4_i1]	3.831	5.62×10^{-10}	3.01×10^{-08}
Glutathione-specific gamma-glutamylcyclotransferase 1 [DN33160_c6_g4_i3]	3.499	4.20×10^{-08}	1.74×10^{-06}
Facilitated trehalose transporter [DN38388_c3_g2_i2]	3.223	5.0×10^{-07}	1.81×10^{-05}
(B) Down-regulated transcripts			
Krasavietz (translational regulator) [DN40711_c3_g1_i6]	−9.075	9.57×10^{-16}	1.07×10^{-13}
Ejaculatory bulb-specific protein 3 [DN40663_c2_g4_i1]	−3.538	9.05×10^{-07}	3.15×10^{-05}
Zinc metalloproteinase [DN40175_c2_g3_i1]	−2.702	3.40×10^{-05}	9.15×10^{-04}
Zinc metalloproteinase [DN40175_c2_g2_i4]	−2.908	1.60×10^{-05}	4.58×10^{-04}
Nucleolin 2 [DN39495_c4_g2_i1]	−2.830	2.36×10^{-05}	6.55×10^{-04}
Nucleolin 2 [DN39495_c4_g2_i2]	−2.752	8.15×10^{-06}	2.45×10^{-04}
Nucleolin 2 [DN35335_c5_g1_i2]	−2.847	1.36×10^{-05}	3.97×10^{-04}
Fatty acid-binding protein [DN39361_c0_g3_i1]	−8.034	2.34×10^{-10}	1.28×10^{-08}
Retinoid-inducible serine carboxypeptidase [DN39186_c2_g1_i4]	−2.791	5.21×10^{-06}	1.61×10^{-04}
Dopamine N-acetyltransferase [DN38477_c5_g1_i2]	−2.571	1.00×10^{-05}	2.96×10^{-04}
Vascular endothelial growth factor C [DN38450_c1_g5_i3]	−2.857	3.05×10^{-06}	9.74×10^{-05}
Ecdysteroid-regulated 16 kDa protein [DN38275_c0_g2_i2]	−2.691	2.45×10^{-05}	6.77×10^{-04}
Sodium-independent sulfate anion transporter [DN37924_c5_g1_i1]	−6.290	9.21×10^{-06}	2.74×10^{-04}
Helicase [DN37874_c0_g1_i2]	−2.695	1.59×10^{-05}	4.57×10^{-04}
Invertebrate-type lysozyme 3 [DN36379_c2_g2_i1]	−9.106	4.16×10^{-13}	3.24×10^{-11}
Protease inhibitor-like [DN37357_c5_g1_i2]	−6.909	2.62×10^{-09}	1.25×10^{-07}
Lysozyme c-1 [DN35555_c0_g1_i2]	−4.140	2.15×10^{-06}	7.05×10^{-05}
Probable cytochrome P450 9f2 [DN35329_c1_g1_i1]	−2.921	9.36×10^{-06}	2.78×10^{-04}
Protein takeout [DN34679_c1_g1_i6]	−3.838	1.64×10^{-07}	6.31×10^{-06}
Sphingomyelin phosphodiesterase [DN34603_c5_g1_i1]	−9.233	2.05×10^{-05}	5.76×10^{-04}
Venom allergen [DN33750_c0_g1_i3]	−5.093	8.37×10^{-10}	4.40×10^{-08}
Synaptic vesicle glycoprotein [DN33233_c5_g1_i10]	−6.485	6.20×10^{-06}	1.89×10^{-04}
Trypsin [DN31717_c0_g2_i1, i2]	−7.446	1.09×10^{-13}	9.35×10^{-12}
Trypsin [DN31717_c0_g2_i2]	−4.595	5.53×10^{-06}	1.70×10^{-04}
Trypsin [DN31717_c0_g1_i1]	−4.957	1.04×10^{-07}	4.12×10^{-06}
Chitinase-like protein Idgf2 [DN31674_c2_g3_i1]	−4.357	2.08×10^{-06}	6.84×10^{-05}

NOTE.—We note that the most differentially expressed genes were uncharacterized proteins with no functional annotation—only genes with putative annotation are shown here in this table. At least 4-fold expression change (logFC values shown) with P-value (P value) cut-off for false discovery rate (FDR) set to 0.001 was required for a gene to be considered differentially expressed. Isoform TMM values equal or larger than 5.0 in at least one of the libraries were required for a gene to be included.

for not only energy metabolism similarly to *Wigglesworthia* in tsetse flies (Michalkova et al. 2014), but proline can be also utilized for peptidoglycan synthesis. Moreover, proline is essential for trypanosome survival in tsetse flies (Mantilla et al. 2017) and could be also involved in triggering host–symbiont interactions as shown for example in *Xenorhabdus/Photorhabdus* symbionts (closely related to *Arsenophonus*) of nematodes (Crawford et al. 2010). Interestingly,

biosynthetic genes for B-vitamins and cofactors were not highly expressed in bacteriocytes of adult insects with the exception of the lipoic acid pathway and several individual enzymes (figs. 2 and 3; supplementary 1, Supplementary Material online). This finding suggests that B-vitamins in adult blood-sucking insects, although essential, do not represent such a prominent case of host–symbiont cooperation as EAAs in plant-sap sucking insects. Such finding is not entirely
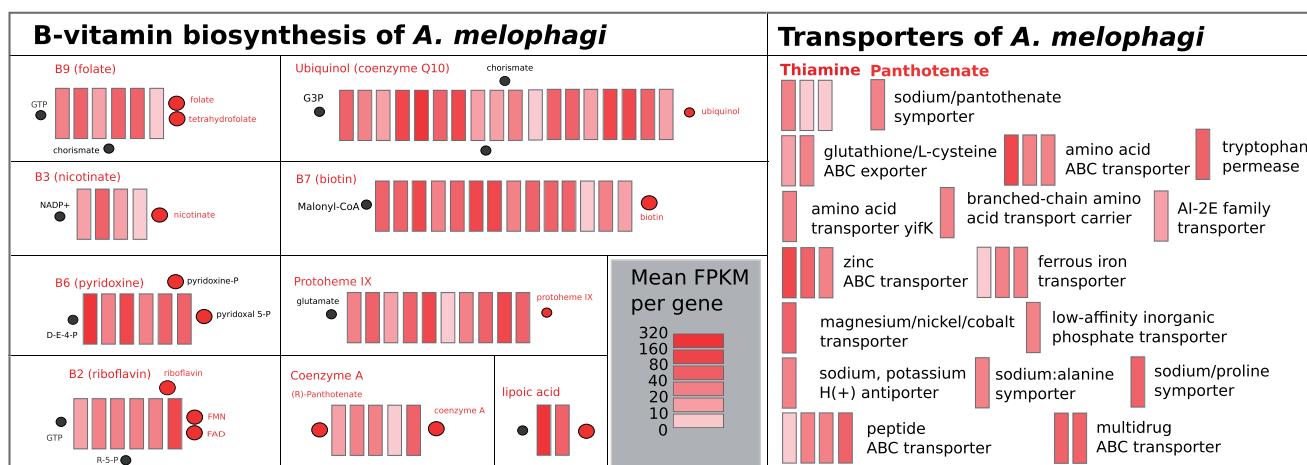
FIG. 3.—Expression of biosynthetic pathways for B-vitamins (B2, B3, B6, B7, B9), cofactors (ubiquinol, protoheme, coenzyme A, and lipoic acid), and various transporters in *Arsenophonus melophagi*. Individual genes are represented by rectangles and filled by heatmap colors representing level of expression (mean FPKM from five replicates).

unexpected because amino acids are crucial for proteosynthesis and growth of the organism, but vitamin-based cofactors are only needed by some of its enzymes. It is, however, likely that these B-vitamins are needed most in particular host life stages such as oocytes and developing larvae that were not included in this study. At least two B-vitamins (thiamine and pantothenate) are acquired by *A. melophagi* from the insect host/blood and it is unknown if all the remaining B-vitamins that are actually synthesized are acquired by the host. An evolutionarily interesting case of a B-vitamin synthesized by symbionts, but probably not provided to the host, was found in cicadas. Obligate cosymbiont of cicadas, *Hodgkinia cicadicola*, retained cobalamine-dependent methionine biosynthesis, so it has to devote at least 7% of its proteome to synthesize cobalamine—vitamin B12 (McCutcheon et al. 2009).

How Specialized Are Midgut Bacteriocytes for Symbiont Maintenance?

We saw only a few host transporters up-regulated specifically in the bacteriocytes of *M. ovinus*, as would be expected in a system without a symbiosomal membrane and with an obligate symbiont still retaining numerous specific and nonspecific transporters (fig. 3; supplementary 2, Supplementary Material online). Because transporters likely import essential nutrients from the gut lumen (and hemolymph) to the endothelial cells along the whole gut, we are unable to fully recognize which of these highly expressed transporters transport the nutrients utilized also by symbionts. Two transporters clearly up-regulated in bacteriocytes (facilitated trehalose transporter and zinc transporters; supplementary 2, Supplementary Material online) support the symbiotic association by providing the metabolites to be processed either by

the host bacteriocyte cell or the symbiont cells. We note that the sodium-dependent multi-vitamin transporter is not significantly up-regulated in bacteriocytes (contrary to the situation in tsetse flies), but rather highly expressed along the entire gut.

Because the bacteriocyte cells are localized in the gut, they likely have to express at least a handful of genes for digestion of the food source and cannot be as specialized as in insects where the bacteriome is freely in the hemolymph. Several hypotheses can be put forward concerning the cell specialization of midgut bacteriocytes. First, these gut cells could be quite polarized and therefore able to carry out a number of different functions depending on their localization. For example, different transporters are likely to be expressed on the gut lumen side compared with the hemolymph side. Our data are averaged from the total tissue and do not allow distinguishing such subtle expression differences that would be only detectable with single-cell RNA-Seq. Second, the bacteriocyte cells are so tightly packed with endosymbionts that they could possibly be no longer able to function as regular midgut cells and could be only functioning as a specialized space for harboring endosymbionts, with only a few essential host genes up-regulated. Third, because the symbiotic system is relatively recent (based on *A. melophagi* genome features and phylogenetic inference) when compared with, for example, the *Buchnera*-aphid system, mechanisms regulating gene expression could be poorly established, relatively random, or compensated by posttranscriptional regulation at both the host and symbiont level. Such differentially regulated genes would be likely invisible by our methodology. Because these three scenarios are not mutually exclusive, we recognize that the outcome is likely a gene/pathway-specific trade off between these and likely a few additional explanations.

Iron and Zinc Regulation and Its Role in Blood Digestion and Symbiosis Maintenance

Iron regulation (i.e., acquisition, utilization, storage, and transport) is an essential function for blood-sucking insects needed to avoid iron toxicity. It is usually carried out by two iron binding proteins (IBP)—ferritin and transferrin. Similar to tsetse flies, heavy and light chain transcripts of ferritin are highly expressed ubiquitously along the whole gut, confirming their general role in iron transport and storage (Strickler-Dinglasan et al. 2006). Particularly, ferritin is involved in sequestering iron from a blood meal to avoid oxidative stress. Transferrin was down-regulated 4.5-fold in bacteriocytes, suggesting that its role is much more specific and perhaps connected to regulation of microorganisms in the gut. It was shown that it can also act as an antimicrobial protein sequestering iron from pathogens (Yoshiga et al. 2001). *Arsenophonus melophagi* does not retain its own ferritin like *W. glossinidia*, but codes and expresses the iron transporter genes *feoABC* and the ferric transcriptional regulator (*fur*).

Surprisingly, multiple lines of evidence support a significance of zinc for this obligate symbiosis. Zinc is generally the second most abundant metal (after iron) in most organisms. It is essential for hundreds of enzymes and zinc finger-containing transcription factors (Coleman 1992; Coleman 1998; McCall et al. 2000). Several transcripts for zinc transporters are highly up-regulated in bacteriocytes and very lowly expressed in the gut (table 2). Furthermore, the zinc ABC transporter of *A. melophagi* symbionts is highly expressed (fig. 3; supplementary 1, Supplementary Material online) and thus supports that zinc is not only imported into the cytoplasm of bacteriocytes, but also imported into *A. melophagi* cells. Host zinc proteases are likely enzymes predominantly using zinc in the gut and two of them were shown to be highly expressed in the gut of tsetse flies (Yan et al. 2002). In our study, many zinc proteases were highly expressed in the rest of the gut, but down-regulated in bacteriocytes (table 2; supplementary 2, Supplementary Material online), suggesting that zinc is mostly used by symbionts in the bacteriome section of the midgut. Why is the host providing zinc specifically to the symbionts? High concentrations of zinc could be toxic to its cells, so it could be using the symbiotic tissue for zinc removal or the symbionts need it at high quantities for an essential function the host is dependent on as well.

A question of particular importance is therefore which *A. melophagi* enzymes rely on zinc. There are numerous zinc-dependent enzyme candidates obvious from our data (supplementary 1, Supplementary Material online), particularly several zinc proteases (TldD, TldE, HtpX, FtsH, RseP, YebA) and a putative metalo-beta-lactamase (GloB). Zinc proteases are involved in proteolysis, suggesting that *A. melophagi* symbionts acquire amino acids by digesting peptide bonds of proteins. This finding could explain why aposymbiotic tsetse flies have difficulties digesting blood (Pais et al. 2008). Beta-lactamases are enzymes that provide bacteria with resistance to beta-lactam antibiotics such as penicillin, ampicillin, and many others. Metalo-beta-lactamases in particular are well-known for their resistance to a broad spectrum of beta-lactam antibiotics and beta-lactamase inhibitors (Bradford 2001; Drawz and Bonomo 2010). Because sheep in the Czech Republic are frequently treated with beta-lactam antibiotics to avoid bacterial infections, it is tempting to speculate that *Arsenophonus* symbionts use their metalo-beta-lactamase (and *Melophagus* supports them by providing zinc) to avoid elimination by antibiotics circulating in sheep blood. If confirmed by additional experiments, it would be, to our knowledge, the first case of an obligate endosymbiont retaining antimicrobial resistance (AMR) genes to survive in its blood-sucking host frequently exposed to antibiotics.

Symbiotic Insect Immunity in the Absence of Peritrophic Matrix

Among the transcripts up-regulated in the whole gut and down-regulated in bacteriocytes (supplementary 2, Supplementary Material online), the antimicrobial peptide attacin (–3.3 FC) and two lysozymes (–550.9 FC and –17.6 FC) are very likely responsible for controlling bacterial infections in the gut and not targeting obligate symbionts *A. melophagi* (fig. 4). Strikingly, PGRP-LB, although highly expressed in both tissues, is not up-regulated in bacteriocytes (fig. 4, supplementary 3, Supplementary Material online) as was reported for the tsetse fly system (Wang et al. 2009; Bing et al. 2017). This finding implies that the host likely unselectively recycles peptidoglycan along the whole gut and, because GNBP has very low expression along the gut, it is unknown if there is an additional mechanism targeting environmental infections in specific sections of the midgut. Possible implications of this difference might be either that the host does not target *S. melophagi* and *B. melophagi*, but is currently "domesticating" these bacteria because they can increase its fitness (e.g., by providing thiamine or pantothenate not synthesized but needed by *Arsenophonus*), or that the host has evolved as partially immunocompromised because of its relatively bacteria-free diet and is therefore unable to recognize the facultative symbionts. The latter explanation would be supported by the lack of peritrofic matrix in *M. ovinus* reported by early microscopy studies (Waterhouse 1953; Lehane 1997). One of essential functions of this non-cellular membrane that is generally found in all other Hippoboscidae species (Waterhouse 1953; Lehane 1997) is providing a barrier to infection by pathogens. Our RNA-Seq data corroborate the microscopy observations because we did not detect any chitin synthesis genes of insect origin expressed in the gut.

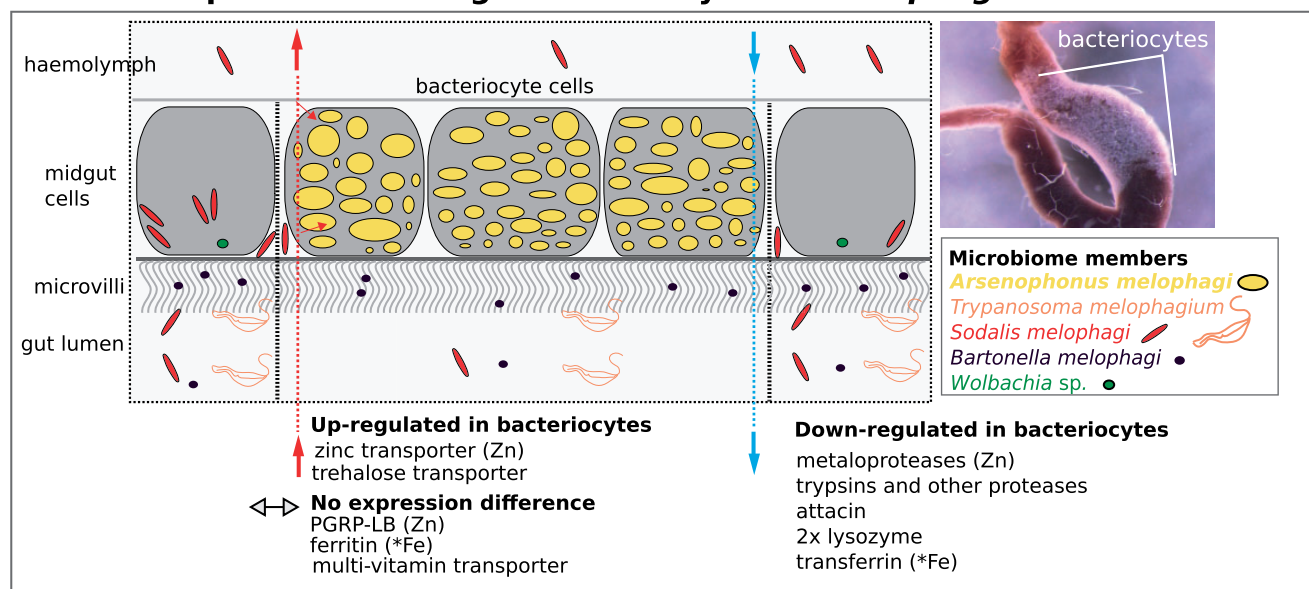## Gene expression in midgut bacteriocytes of *Melophagus ovinus*

Fig. 4.—Schematic reconstruction of nutritional and immunity processes revealed as involved in host–symbiont interactions in the *Melophagus ovinus* midgut cells harboring *Arsenophonus* symbionts. Selected host genes of interest up-regulated or down-regulated in bacteriocytes are highlighted. Zinc-dependent enzymes are denoted by Zn, iron-binding proteins are denoted by Fe. Localization of the midgut bacteriocytes is shown on a small inset figure.

Gene Duplications and Novel Genes Are Recruited for the Symbiosis Maintenance

As usual in insect RNA-Seq studies, many differentially expressed and/or highly expressed genes observed in our data (supplementary 2, Supplementary Material online) code conserved hypothetical proteins, orphan proteins with no homology to proteins in protein databases, and numerous gene duplications. These genes need to be examined experimentally in future studies. For example, RNA-Seq study on aphid bacteriocytes and whole mount in situ hybridizations of overrepresented transcripts encoding aphid-specific orphan proteins has revealed a novel family of small cysteine-rich proteins with signal peptides (Shigenobu and Stern 2013). At least some orphan genes in the pea aphid genome thus likely (International Aphid Genomics Consortium 2010) evolved to assist in lineage-specific traits, such as symbiosis. The role of gene and isoform specialization and recruitment for functioning in bacteriocytes was observed also in our data, for example in zinc transporters and proteinases (supplementary 2, Supplementary Material online), implying that it is a more general mechanism occurring outside of sap-feeding insects.

Host–Symbiont Interactions in Blood-Sucking and Plant-Sap Sucking Insects—Similarities and Differences

Unlike in plant-sap sucking insects where the host is strongly involved in the EAA biosynthetic pathways (sometimes even forming mosaic pathways requiring genes from both the host and its symbiont), we found no evidence for such intimate interactions in our *M. ovinus* data for any biosynthetic pathways (table 1; fig. 4; supplementary 2, Supplementary Material online). Whether this is a common situation in all blood-sucking insects, or if it applies only to hosts with symbionts having medium-sized genomes, remains to be investigated. As insect symbionts are either present directly in the bacteriocyte cytoplasm (common in blood-sucking and omnivorous), or enclosed in a so-called symbiosomal membrane (common in plant-sap sucking insects), the cellular localization of symbionts may also be responsible for the intimacy of the relationships. The cytoplasm-harbored symbionts, such as *A. melophagi* in *M. ovinus*, have direct access to all nutrients available in the host cells. For symbionts covered by a symbiosomal membrane, their host controls which nutrients will be available to the symbionts by adjusting the expression of transporters through the symbiosomal membrane localization (Price et al. 2014; Duncan et al. 2014). This situation then inevitably leads to highly interconnected relationships (and accelerates symbiont genome reduction). Evaluating the exact impact of this less intimate symbiotic integration is much needed to understand different factors driving genome reduction in different endosymbiont lineages.

Conclusions

Using RNA-Seq with five biological replicates of the midgut cells housing bacteria and the rest of the gut, we uncovered the interactions between a blood-sucking parasite and its

obligate endosymbiont. We found strong evidence for importance of zinc in the system possibly caused by participation of symbionts on blood-digestion; and for different immunity mechanisms controlling symbionts than in closely related tsetse flies. Our results show that adults of this blood-sucking insect are much less intimately involved in cooperation on biosynthesis of nutrients than plant-sap sucking insects. This finding is likely interconnected to several features observed in symbionts in blood-sucking arthropods, particularly intracytoplasmic midgut localization of bacteria, their less severe genome reduction, and younger associations caused by frequent evolutionary losses and replacements. Because the specialized bacteriome section of *M. ovinus* is clearly distinct from the rest of the midgut based on both morphology and gene expression, it will be interesting to compare gene expression of this specialized tissue to other blood-sucking insects in future.

Supplementary Material

Supplementary data are available at FigShare (https://doi.org/10.6084/m9.figshare.6146777).

Acknowledgments

Author Contributions

F.H. designed and coordinated the study, carried out the preparation of samples, analyses of data, and drafted the manuscript. V.H. carried out analyses of immune genes and participated in the study design and manuscript revisions. A.D. designed the study and participated in manuscript revisions. All authors read and approved the final manuscript.

Literature Cited

Aksoy S. 1995. Molecular analysis of the endosymbionts of tsetse flies: 16S rDNA locus and over-expression of a chaperonin. Insect Mol Biol. 4(1):23–29.

Anselme C, Vallier A, Balmand S, Fauvarque MO, Heddi A. 2006. Host PGRP gene expression and bacterial release in endosymbiosis of the weevil Sitophilus zeamais. Appl Environ Microbiol. 72(10):6766–6772.

Anselme C, et al. 2008. Identification of the Weevil immune genes and their expression in the bacteriome tissue. BMC Biol. 6(1):43.

Arrese EL and Soulages JL. 2010. Insect fat body: energy, metabolism, and regulation. Annu Rev Entomol. 55:207–225.

Baumann P, Baumann L, Clark MA. 1996. Levels of Buchnera aphidicola chaperonin GroEL during growth of the aphid Schizaphis graminum. Curr Microbiol. 32(5):279–285.

Belda E, Silva FJ, Pereto J, Moya A. 2012. Metabolic networks of Sodalis glossinidius: a systems biology approach to reductive evolution. PLoS One 7(1):e30652.

Bing X, et al. 2017. Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. Proc R Soc B. 284(1857):20170360.

Boyd BM, et al. 2016. Two bacterial genera, Sodalis and Rickettsia, associated with the seal louse Proechinophthirus fluctus (Phthiraptera: Anoplura). Appl Environ Microbiol. 82(11):3185–3197.

Bradford PA. 2001. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev. 14(4):933–951.

Brucker RM, Funkhouser LJ, Setia S, Pauly R, Bordenstein SR. 2012. Insect Innate Immunity Database (IIID): an annotation tool for identifying immune genes in insect genomes. PLoS One 7(9):e45125.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28(4):464–469.

Charles H, et al. 2011. A genomic reappraisal of symbiotic function in the aphid/Buchnera symbiosis: reduced transporter sets and variable membrane organisations. PLoS One 6(12):e29096.

Chrudimský T, Husník F, Nováková E, Hypša V. 2012. Candidatus Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont Sodalis glossinidius. PLoS One 7(7):e40354.

Coleman JE. 1992. Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. Annu Rev Biochem. 61(1):897–946.

Coleman JE. 1998. Zinc enzymes. Curr Opin Chem Biol. 2(2):222–234.

Crawford JM, Kontnik R, Clardy J. 2010. Regulating alternative lifestyles in entomopathogenic bacteria. Curr Biol. 20(1):69–74.

Douglas AE. 2016. How multi-partner endosymbioses function. Nat Rev Microbiol. 14(12):731–743.

Drawz SM, Bonomo RA. 2010. Three decades of beta-lactamase inhibitors. Clin Microbiol Rev. 23(1):160–201.

Duncan RP, et al. 2014. Dynamic recruitment of amino acid transporters to the insect/symbiont interface. Mol Ecol. 23(6):1608–1623.

Eichler S, Schaub GA. 2002. Development of symbionts in triatomine bugs and the effects of infections with trypanosomatids. Exp Parasitol. 100(1):17–27.

Fan Y, Thompson JW, Dubois LG, Moseley MA, Wernegreen JJ. 2013. Proteomic analysis of an unculturable bacterial endosymbiont (Blochmannia) reveals high abundance of chaperonins and biosynthetic enzymes. J Proteome Res. 12(2):704–718.

Fischer JD, Holliday GL, Thornton JM. 2010. The CoFactor database: organic cofactors in enzyme catalysis. Bioinformatics 26(19):2496–2497.

Fukatsu T, Ishikawa H. 1993. Occurrence of chaperonin 60 and chaperonin 10 in primary and secondary bacterial symbionts of aphids: implications for the evolution of an endosymbiotic system in aphids. J Mol Evol. 36(6):568–577.

Gerardo NM, Wilson AC. 2011. The power of paired genomes. Mol Ecol. 20(10):2038–2040.

Gerardo NM, et al. 2010. Immunity and other defenses in pea aphids, Acyrthosiphon pisum. Genome Biol. 11(2):R21.

Gonzalez-Domenech CM, et al. 2012. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two Blattabacterium cuenoti strains, primary endosymbionts of cockroaches. BMC Microbiol. 12:S5.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.

Gruber-Vodicka HR, Seah BK, Pruesse E. 2019. phyloFlash—rapid SSU rRNA profiling and targeted assembly from metagenomes. bioRxiv 521922.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. Proc Natl Acad Sci USA. 108(7):2849–2854.

Hansen AK, Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. Mol Ecol. 23(6):1473–1496.

Husnik F. 2018. Host–symbiont–pathogen interactions in blood-feeding parasites: nutrition, immune cross-talk and gene exchange. Parasitology 145(10):1294–1303.

Husnik F, McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. Proc Natl Acad Sci USA. 113(37):E5416–E5424.

Husnik F, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153(7):1567–1578.

International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8:e1000313.

Karp PD, et al. 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform. 11(1):40–79.

Kirkness EF, 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 USA. 107(27):12168–12173.

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Research 6:1287.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357–359.

Lehane MJ. 1997. Peritrophic matrix structure and function. Annu Rev Entomol. 42(1):525–550.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12(1):323.

Login FH, et al. 2011. Antimicrobial peptides keep insect endosymbionts under control. Science (80-) 334(6054):362–365.

Lu Hl, Chang Cc, Wilson ACC. 2016. Amino acid transporters implicated in endocytosis of *Buchnera* during symbiont transmission in the pea aphid. Evodevo 7:24.

Luan J-B, et al. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. Genome Biol Evol. 7(9):2635–2647.

Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. 2012. The central role of the host cell in symbiotic nitrogen metabolism. Proc R Soc B. 279(1740):2965–2973.

MacDonald SJ, Thomas GH, Douglas AE. 2011. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. Mol Ecol. 20(10):2073–2084.

Mantilla BS, et al. 2017. Proline metabolism is essential for *Trypanosoma brucei brucei* survival in the tsetse vector. PLoS Pathog. 13(1):e1006158–29.

Manzano-Marín A, Simon J-C, Latorre A. 2016. Reinventing the wheel and making it round again: evolutionary convergence in *Buchnera*—*Serratia* symbiotic consortia between the distantly related Lachninae Aphids *Tuberolachnus salignus* and *Cinara cedri*. Genome Biol Evol. 8(5):1440–1458.

Mao M, Yang X, Bennett GM. 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. Proc Natl Acad Sci USA. 115(50):E11691–E11700.

McCall KA, Huang C, Fierke CA. 2000. Function and mechanism of zinc metalloenzymes. J Nutr. 130(5):1437S–1446S.

McClure R, et al. 2013. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. 41(14):e140.

McCutcheon J, McDonald B, Moran N. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci USA. 106(36):15394–15399.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 10(1):13–26.

Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. 2014. Obligate symbiont-generated vitamin B6 is critical to maintain proline

homeostasis and fecundity in tsetse flies. Appl Environ Microbiol. 80(18):5844–5853.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. Annu Rev Microbiol. 68(1):195–215.

Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 42(1):165–190.

Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. Curr Biol. 24(14):R640–641.

Nakabachi A, Ishikawa H. 1999. Provision of riboflavin to the host aphid, *Acyrthosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. J Insect Physiol. 45(1):1–6.

Nakabachi A, et al. 2005. Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. Proc Natl Acad Sci USA. 102(15):5477–5482.

Nikoh N, et al. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc Natl Acad Sci USA. 111(28):10257–10262.

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(18):6189–6199.

Nováková E, Hypša V, Nguyen P, Husník F, Darby AC. 2016. Genome sequence of *Candidatus* Arsenophonus lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). Stand Genomic Sci. 11: 72.

Oakeson KF, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. Genome Biol Evol. 6(1):76–93.

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(19):5965–5974.

Poliakov A, et al. 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol Cell Proteomics. 10(6):M110.007039.

Price DRG, et al. 2014. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. Proc Natl Acad Sci USA. 111(1):320–325.

Ratzka C, Gross R, Feldhaar H. 2013. Gene expression analysis of the endosymbiont-bearing midgut tissue during ontogeny of the carpenter ant *Camponotus floridanus*. J Insect Physiol. 59(6):611–623.

Říhová J, Novaková E, Husník F, Hypša V. 2017. *Legionella* becoming a mutualist: adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. Genome Biol Evol. 9(11):2946–2957.

Rio RV, et al. 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. MBio 3(1):1–13.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140.

Roje S. 2007. Vitamin B biosynthesis in plants. Phytochemistry 68(14):1904–1921.

Shigenobu S, Stern DL. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. Proc R Soc B. 280(1750):20121952.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.

Sloan DB, et al. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol Biol Evol. 31(4):857–871.

Small RW. 2005. A review of *Melophagus ovinus* (L.), the sheep ked. Vet Parasitol. 130(1–2):141–155.

Strickler-Dinglasan PM, Guz N, Attardo G, Aksoy S. 2006. Molecular characterization of iron binding proteins from *Glossina morsitans morsitans* (Diptera: Glossinidae). Insect Biochem Mol Biol. 36(12):921–933.

Thomas GH, et al. 2009. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. BMC Syst Biol. 3:24.

Van Leuven JT, Meister RC, Simon C, McCutcheon JP. 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. Cell 158:1270–1280.

Vigneron A, et al. 2014. Insects recycle endosymbionts when the benefit is over. Curr Biol. 24(19):2267–2273.

Wang JW, Aksoy S. 2012. PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. Proc Natl Acad Sci USA. 109(26):10552–10557.

Wang JW, Wu YN, Yang GX, Aksoy S. 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. Proc Natl Acad Sci USA. 106(29):12133–12138.

Waterhouse DF. 1953. The occurrence and significance of the peritrophic membrane, with special reference to adult lepidoptera and diptera. Aust J Zool. 1(3):299–318.

Weiss BL, Wang J, Aksoy S. 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol. 9(5):e1000619.

Yan J, Cheng Q, Li CB, Aksoy S. 2002. Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. Insect Mol Biol. 11(1):57–65.

Yoshiga T, et al. 2001. *Drosophila melanogaster* transferrin—Cloning, deduced protein sequence, expression during the life cycle, gene localization and up-regulation on bacterial infection. Eur J Biochem. 260(2):414–420.

Associate editor: Richard TE Cordaux