

# Insect—symbiont gene expression in the midgut bacteriocytes of a blood-sucking parasite

Husnik Filip<sup>1,2,‡</sup>, Hypsa Vaclav<sup>1,2</sup>, and Darby Alistair<sup>3</sup>

<sup>1</sup> Biology Centre AS CR, Institute of Parasitology, Branišovská 31, České Budějovice 37005, Czech Republic

<sup>2</sup> University of South Bohemia, Faculty of Science, Branišovská 31, České Budějovice 37005, Czech Republic

<sup>3</sup> University of Liverpool, Institute of Integrative Biology, Crown Street, Liverpool L69 7ZB, UK

Corresponding authors: Filip Husnik<sup>‡</sup> [[filip.husnik@gmail.com](mailto:filip.husnik@gmail.com)] and Alistair Darby [[acdarby@liv.ac.uk](mailto:acdarby@liv.ac.uk)]

## Abstract

**Background.** Animals interact with a diverse array of both beneficial and detrimental microorganisms. These interactions sometimes spark obligate symbioses where the host depends on beneficial bacteria for survival and reproduction. In insects, these obligate symbioses in many cases allow feeding on nutritionally unbalanced diets such as plant sap and vertebrate blood. It is, however, still not clear how are these obligate intracellular 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.

**Results.** Here, we analyzed interactions at the symbiotic interfaces of an obligately blood-sucking parasite of sheep, the louse fly *Melophagus ovinus*. We assembled a reference transcriptome from one male and one female individual and used RNA-Seq with five biological replicates to compare expression in the midgut cells housing bacteria to the rest of the gut (foregut-hindgut). We focused on nutritional and immunity interactions between the insect host and its obligate symbiont *Arsenophonus melophagi*, and also generated lower-coverage data for three facultative bacterial symbionts (*Sodalis melophagi*, *Bartonella melophagi*, and *Wolbachia* sp.) and one facultative eukaryote *Trypanosoma melophagium*. 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 likely different immunity mechanisms controlling the symbionts than in closely related tsetse flies.

**Conclusions.** 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 some (but not all) plant-sap sucking insects such as aphids, psyllids, whiteflies, and mealybugs. This finding is likely interconnected to several features observed in symbionts in blood-sucking arthropods, particularly their midgut intracellular localization (as opposed to being localized in truly specialized bacteriocytes), intracytoplasmic presence (as opposed to having an outermost host-derived ‘symbiosomal’ membrane), less severe genome reduction, and relatively recent associations caused by frequent evolutionary losses and replacements.

**Keywords:** RNA-Seq, B-vitamins, parasites, symbiotic bacteria, interactions, zinc, immunity

**Data deposition:** Raw RNA-Seq data were made available through the European Nucleotide Archive (ENA) database under the study accession number PRJEB30632. All assemblies and additional large supplementary files are available on FigShare [<https://doi.org/10.6084/m9.figshare.6146777.v1>]. All commands used for data analyses are available on Github [<https://github.com/filip-husnik/melophagus>].

---

‡ Present address: University of British Columbia, Faculty of Science, Vancouver, British Columbia, Canada

## Background

Nutritional supplementation from symbiotic bacteria allowed several insect groups to become specialized on nutritionally unbalanced diets such as vertebrate blood and plant sap. Nutrients so far shown to be provided by bacteria include essential amino-acids and some B-vitamins in plant-sap sucking hosts, B-vitamins in blood-sucking hosts, and symbiotic bacteria (or protists and fungi) are also involved in nitrogen recycling in nitrogen-restricted environments such as in wood-feeding insects (Moran et al. 2008, McCutcheon and Moran 2011, 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 sucking Hemiptera vs. Diptera) or the symbiont phylogeny (e.g. Gammaproteobacteria vs. Bacteroidetes)? Since the small symbiont genomes are usually subsets of well-known bacterial genomes such as *E. coli* and code only a handful of unknown genes, it is feasible to roughly predict their metabolism either by simple mapping of present genes on metabolic pathways (Hansen and Moran 2013, 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 non-model 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 by the GS/GOGAT (glutamine synthase/glutamine oxoglutarate aminotransferase) cycle. Flux-balance analyses and additional data (Macdonald et al. 2012) suggest that waste ammonia is recycled predominantly by the host cell

(bacteriocyte) and that aphid aminotransferases (ornithine AT, branched-chain AT, and aspartate AT) incorporate ammonia-derived nitrogen into carbon skeletons synthesized by *Buchnera* to generate EAAs. These hypotheses were also verified experimentally (Macdonald et al. 2012). 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 2013), 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, e.g. biotin, riboflavin, and folate (Hansen and Moran 2013, Moran and Bennett 2014). It is unclear which B-vitamins are only used by symbionts and which are in addition also provided to their hosts. The only piece of experimental evidence implies that young symbiotic aphids are provided with riboflavin by their *Buchnera* endosymbionts (Nakabachi and Ishikawa 1999).

Host-symbiont cooperation is exceptionally dependent on the well-working 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). Host amino acid transporters can be involved in symbiont maintenance – some of them were extensively duplicated and specialized for bacteriocyte transfer (Duncan et al. 2014) and symbiont control (Price et al. 2013, 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, pantothenate 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 arthropods either somehow efficiently extract rare nutrients from their blood diet or rely on extracellular gut bacteria acquired from the environment, e.g. 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 (e.g. facultative vs. obligate dependence on the blood diet), association with the host (temporary vs. permanent), host species (e.g. mammals vs. birds), enzymatic dependence (e.g. using alternative enzymes not depending on a particular co-factor), 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 and other cofactors such as thiamine in human lice and louse flies are actually transported to the symbiont cells from the blood-diet with a thiamine ABC transporter encoded on the symbiont genome (reviewed in Husnik 2018).

In comparison to the very intimate host-symbiont cooperation on biosynthesis of amino acids in plant-feeding insects, 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 regulate and maintain the symbiosis, particularly a multi-vitamin transporter is up-regulated to shuttle B-vitamins from midgut bacteriocytes to haemolymph 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 2011). In the latter case, they are often engulfed by a host-derived symbiosomal membrane (McCutcheon and

Moran 2011), 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 as in lice symbionts, 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, 2008, Wang et al. 2009, Weiss et al. 2011, Wang and Aksoy 2012, Ratzka et al. 2013, Bing et al. 2017). A cautionary tale showing that insect symbioses are never-ending power struggles rather than idyllic interactions was reported from *Sitophilus* weevils and their recently ‘domesticated’ *Sodalis pierantonius* symbiont (Oakeson et al. 2014). A single antimicrobial peptide (Coleopteracin A) was shown to keep the symbionts under control, and if it was silenced by RNA interference, the symbionts were escaping from the bacteriocytes and spreading into host tissues (Login et al. 2011). The functional role of the symbionts is likely to synthesize a single non-essential amino acid, tyrosin, needed for cuticle hardness and when this benefit is no longer needed, 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.

Here, we use comprehensive RNA-Seq data for an obligately blood-sucking insect harboring several endosymbionts and pathogens to understand metabolic interdependence and host-symbiont interactions. We generated data for an ectoparasite of sheep, *Melophagus ovinus*, and analyzed host-symbiont interactions within its microbiome containing not only the obligate symbiotic bacterium *Arsenophonus melophagi*, but also 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, 2016).

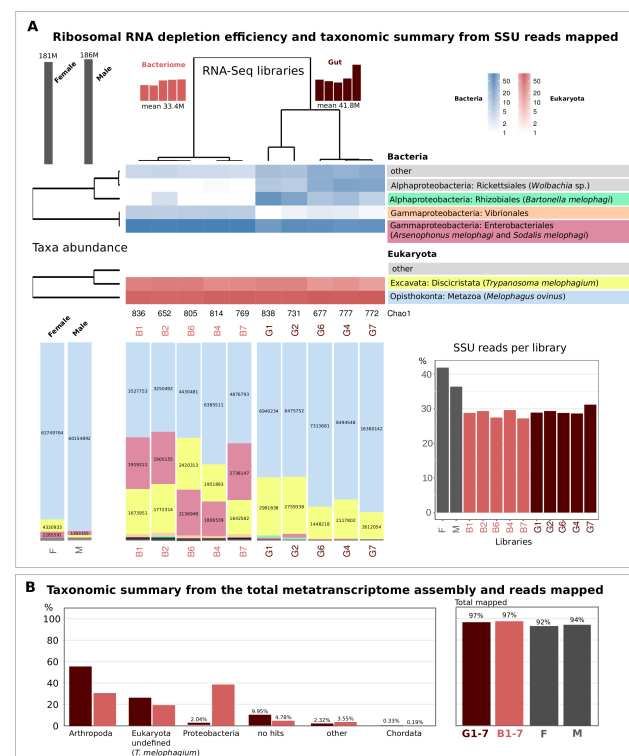
## Results

### Contamination and rRNA depletion efficiency assessment

Since 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 suggesting that the rRNA depletion method used had relatively poor efficiency. 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 expression analyses for *M. ovinus* and *A. melophagi*. Species composition profiles allowed by SSU mapping revealed very homogenous microbiome composition across all samples (Figure 1A) with majority of data from the host *M. ovinus* (and low coverage sheep contamination from blood), its two most prominent bacterial symbionts *Arsenophonus* and *Sodalis melophagi*, and a eukaryotic commensal/pathogen *Trypanosoma melophagium*. Additional lower abundance taxa of interest include for example *Bartonella 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 (Figure 1B).

### Evaluation of symbiont expression

FPKM values obtained by mapping raw data on the *A. melophagi* genome are presented in Suppl. 1 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.



**Fig1 Ribosomal RNA depletion efficiency and species composition/contamination assessment. (A)** Small ribosomal rRNA genes (both bacterial and eukaryotic) and **(B)** Total meta-transcriptome data. 'Chao' values represent total SSU diversity per library.

### 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 *Trypanosoma 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).

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 $\beta$	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 $\beta'$	K	336
rpsM	30S ribosomal protein S13	J	332
rpsU	30S ribosomal protein S21	J	330
rpoA	RNA polymerase subunit $\alpha$	K	327
sodB	Superoxide dismutase [Fe]	P	325

**Tab1: Thirty most highly expressed *A. melophagi* genes.** COG functional categories: [M] Cell wall/membrane/envelope biogenesis; [O] Post-translational 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.

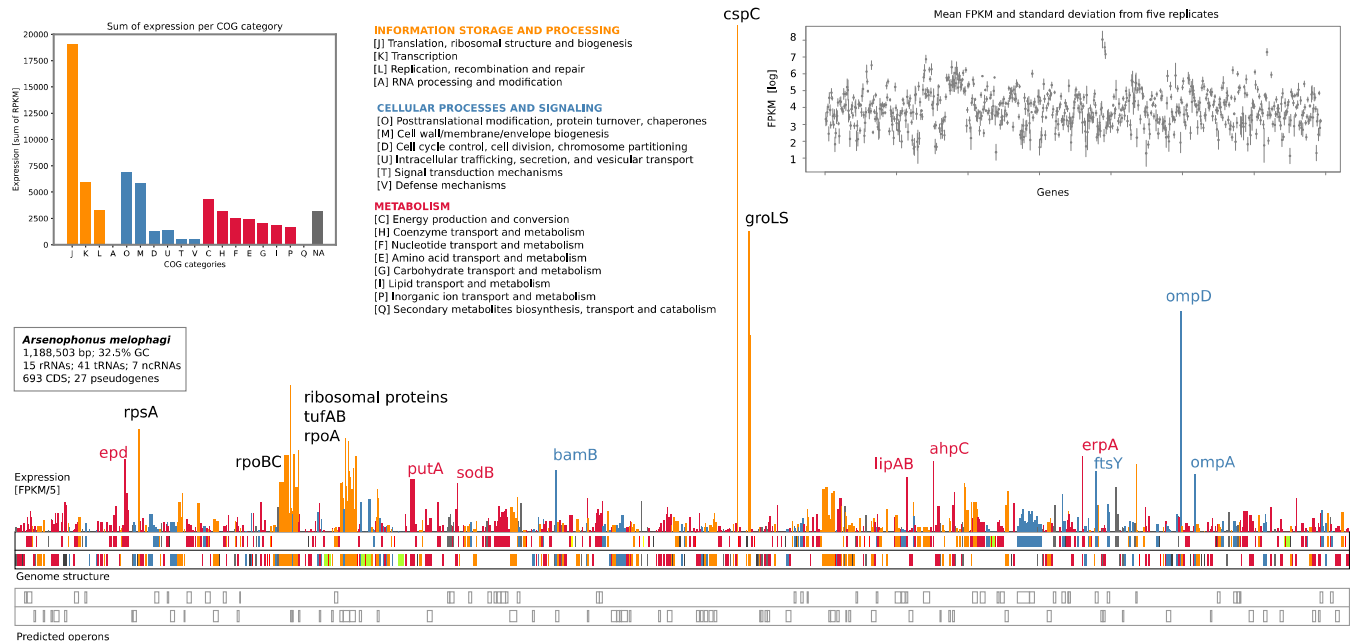
Raw counts of mapped reads, normalized digital expression values, and results of differential expression analyses from EdgeR for gut and bacteriome tissues are available in Suppl. 2. In total, 1353 genes were found significantly up-regulated and 266 genes significantly down-regulated in the

bacteriome section of the midgut (Suppl. 2). We note that the much larger number of up-regulated genes is an overestimate caused by bacterial transcripts that were not excluded from the total assembly for the differential expression step. 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 (*W. pipientis* and *B. melophagi*) promote the viral infection.

#### *Reconstruction of selected pathways involved in host—symbiont—pathogen interactions*

B-vitamin and co-factor metabolism of all members of the microbiome set against the host pathways uncovered utilization of B-vitamins in the gut, acquisition of some B-vitamins (thiamine and pantothenate) by *A. melophagi* from bacteriocytes, but also likely indirect acquisition of other B-vitamins by the host bacteriocytes (e.g. after *A. melophagi* cell lysis). Possible exploitation of this resource by facultative members of the microbiome, i.e. *B. melophagi*, *S. melophagi*, *Wolbachia* sp., and *T. melophagium* is also likely (Fig 3B). However, the functional role of *T. melophagium* was not investigated in much detail as it likely does not regulate expression similarly to other trypanosomatids and it seems to be found along the entire gut (Fig 1). Gene expression data for *Sodalis melophagi* and *Bartonella melophagi* were of relatively low abundance, so 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 3A, Suppl 4). Gene/transcript information used for the reconstruction of the metabolic pathways and immune system can be found in Suppl. 1 and 2.



**Fig2 Linear genome map of *A. 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 since it was depleted and is likely biased. Expression of tRNAs and other short ncRNAs is not shown since these RNAs are under-represented 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.

### Transporters up-regulated in bacteriocytes

Two host transporters possibly involved in maintenance of the symbiotic system were found to be significantly up-regulated in bacteriocytes (Tab 2). These transporters (facilitated trehalose transporter and zinc transporter) regulate 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 e.g. 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 the 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. Several transporters of *T. melophagium* were also evaluated as highly expressed in the bacteriome and/or gut sections (Suppl. 2). Trypanosomes thus likely scavenge available nutrients similarly as the remaining members of the microbiome: *S. melophagi*, *B. melophagi*, and *Wolbachia* sp. (Suppl. 2).

### Additional transcripts of bacterial origin

All transcripts recognized by our cut-off values as of bacterial origin are available in Suppl. 2. 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 (Suppl. 2). 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). Since our reference genome assembly of *M. ovinus* (to double-check HGT candidates) is of extremely low quality, we do not discuss these genes of possible bacterial origin further. Genome data of much higher quality for both the host and its facultative microbiome members would be needed to reliably

distinguish horizontal gene transfer events in the *Melophagus ovinus* genome and microbiome.

## Discussion

**Symbiont expression, B-vitamin biosynthesis, and host-symbiont metabolite exchange.** The expression profile of *A. melophagi* (Fig 2; Suppl. 1; Tab 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 anti-sense ncRNA) with transcription anti-terminator activity. Other highly expressed genes were found to be for heat shock proteins (GroEL and GroES), outer membrane proteins, and porins allowing various small solutes to cross the outer membrane (OmpD, OmpA, BamABDE), enzymes protecting against reactive oxygen species (alkyl hydroperoxide reductase and superoxide dismutase), and iron-sulfur cluster insertion protein ErpA; Fig 2; Suppl. 1). 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 and its subsequent conversion to D-glutamate by the MurI enzyme. Considering that proline is almost always the most common amino acid in insect haemolymph (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 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, B-vitamin and co-factor biosynthetic genes were not highly expressed in bacteriocytes with the exception of the lipoic acid

pathway and several individual enzymes (Fig 2; Fig 3; Suppl. 1). This finding suggests that B-vitamins in blood-sucking insects, although essential, do not represent such a prominent case of host-symbiont cooperation as essential amino acids (EAAs) in plant-sap sucking insects. Such finding is not entirely unexpected since amino acids are crucial for proteosynthesis and growth of the organism, but vitamin-based co-factors are only needed by some of its enzymes. B-vitamins are based on our expression data likely used in relatively low volumes in adults and potentially needed most in particular host life stages that were not studied here (such as oocytes and developing larvae). At least two B-vitamins (thiamine and pantothenate) are acquired by *A. melophagi* from the 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 B-vitamin biosynthesis needed for symbionts, but probably not provided to the host, was found in cicadas. Obligate co-symbiont 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). Future experimental studies will need to be conducted to determine whether a similar scenario also applies to some pathways in blood-sucking insects.

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 a decent set of both specific and non-specific transporters (Fig 3; Suppl. 2). Because transporters likely import essential nutrients from the gut lumen 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; Suppl. 2) 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.

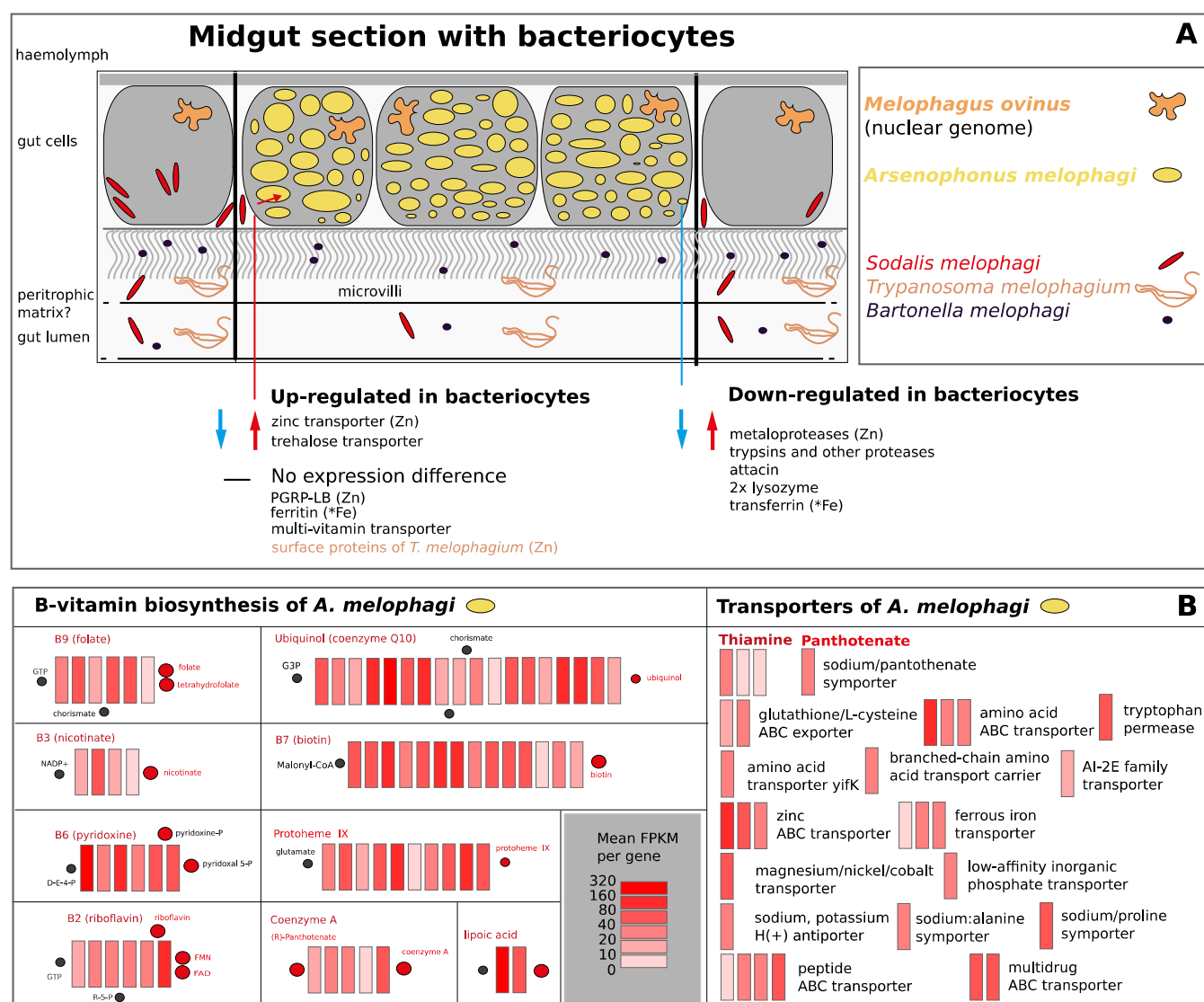
(A) Up-regulated transcripts		Fold change		
Annotation [transcript ID]	logFC	Pvalue	FDR	
FLYWCH-type zinc finger-containing protein 1 [DN33516_c0_g1_i7,i2,i5]	5.792	1.86x10 <sup>-12</sup>	1.32x10 <sup>-10</sup>	
	4.708	4.07x10 <sup>-09</sup>	1.90x10 <sup>-07</sup>	
	3.941	1.60x10 <sup>-09</sup>	8.01x10 <sup>-08</sup>	
tyrosine-protein phosphatase 10D [DN35320_c1_g6_i1]	5.325	1.12x10 <sup>-16</sup>	1.42x10 <sup>-14</sup>	
Paramyosin, short form isoform [DN33516_c0_g1_i2]	4.645	1.30x10 <sup>-10</sup>	7.38x10 <sup>-09</sup>	
zinc transporter foi [DN36258_c3_g8_i2]	4.082	4.90x10 <sup>-10</sup>	2.64x10 <sup>-08</sup>	
[DN37931_c0_g1_i2]	3.430	6.74x10 <sup>-10</sup>	3.58x10 <sup>-08</sup>	
glutathione-specific gamma-glutamylcyclotransferase 1 [DN33160_c6_g4_i1, i3]	3.831	5.62x10 <sup>-10</sup>	3.01x10 <sup>-08</sup>	
	3.499	4.20x10 <sup>-08</sup>	1.74x10 <sup>-06</sup>	
facilitated trehalose transporter [DN38388_c3_g2_i2]	3.223	5.0x10 <sup>-07</sup>	1.81x10 <sup>-05</sup>	

(B) Down-regulated transcripts		Fold change		
Annotation [transcript ID]	logFC	Pvalue	FDR	
krasavietz (translational regulator) [DN40711_c3_g1_i6]	-9.075	9.57x10 <sup>-16</sup>	1.07x10 <sup>-13</sup>	
ejaculatory bulb-specific protein 3 [DN40663_c2_g4_i1]	-3.538	9.05x10 <sup>-07</sup>	3.15x10 <sup>-05</sup>	
zinc metalloproteinase [DN40175_c2_g3_i1, g2_i4]	-2.702	3.40x10 <sup>-05</sup>	9.15x10 <sup>-04</sup>	
	-2.908	1.60x10 <sup>-05</sup>	4.58x10 <sup>-04</sup>	
Nucleolin 2 [DN39495_c4_g2_i1, i2]	-2.830	2.36x10 <sup>-05</sup>	6.55x10 <sup>-04</sup>	
[DN35335_c5_g1_i2]	-2.752	8.15x10 <sup>-06</sup>	2.45x10 <sup>-04</sup>	
	-2.847	1.36x10 <sup>-05</sup>	3.97x10 <sup>-04</sup>	
Fatty acid-binding protein [DN39361_c0_g3_i1]	-8.034	2.34x10 <sup>-10</sup>	1.28x10 <sup>-08</sup>	
Retinoid-inducible serine carboxypeptidase [DN39186_c2_g1_i4]	-2.791	5.21x10 <sup>-06</sup>	1.61x10 <sup>-04</sup>	
Dopamine N-acetyltransferase [DN38477_c5_g1_i2]	-2.571	1.00x10 <sup>-05</sup>	2.96x10 <sup>-04</sup>	
vascular endothelial growth factor C [DN38450_c1_g5_i3]	-2.857	3.05x10 <sup>-06</sup>	9.74x10 <sup>-05</sup>	
ecdysteroid-regulated 16 kDa protein [DN38275_c0_g2_i2]	-2.691	2.45x10 <sup>-05</sup>	6.77x10 <sup>-04</sup>	
sodium-independent sulfate anion transporter [DN37924_c5_g1_i1]	-6.290	9.21x10 <sup>-06</sup>	2.74x10 <sup>-04</sup>	
helicase [DN37874_c0_g1_i2]	-2.695	1.59x10 <sup>-05</sup>	4.57x10 <sup>-04</sup>	
invertase-type lysozyme 3 [DN36379_c2_g2_i1]	-9.106	4.16x10 <sup>-13</sup>	3.24x10 <sup>-11</sup>	
protease inhibitor-like [DN37357_c5_g1_i2]	-6.909	2.62x10 <sup>-09</sup>	1.25x10 <sup>-07</sup>	
lysozyme c-1 [DN35555_c0_g1_i2]	-4.140	2.15x10 <sup>-06</sup>	7.05x10 <sup>-05</sup>	
probable cytochrome P450 9f2 [DN35329_c1_g1_i1]	-2.921	9.36x10 <sup>-06</sup>	2.78x10 <sup>-04</sup>	
protein takeout [DN34679_c1_g1_i6]	-3.838	1.64x10 <sup>-07</sup>	6.31x10 <sup>-06</sup>	
sphingomyelin phosphodiesterase [DN34603_c5_g1_i1]	-9.233	2.05x10 <sup>-05</sup>	5.76x10 <sup>-04</sup>	
venom allergen [DN33750_c0_g1_i3]	-5.093	8.37x10 <sup>-10</sup>	4.40x10 <sup>-08</sup>	
synaptic vesicle glycoprotein [DN33233_c5_g1_i10]	-6.485	6.20x10 <sup>-06</sup>	1.89x10 <sup>-04</sup>	
Trypsin [DN31717_c0_g2_i1, i2]	-7.446	1.09x10 <sup>-13</sup>	9.35x10 <sup>-12</sup>	
[DN31717_c0_g1_i1]	-4.595	5.53x10 <sup>-06</sup>	1.70x10 <sup>-04</sup>	
	-4.957	1.04x10 <sup>-07</sup>	4.12x10 <sup>-06</sup>	
chitinase-like protein Idgf2 [DN31674_c2_g3_i1]	-4.357	2.08x10 <sup>-06</sup>	6.84x10 <sup>-05</sup>	

**Tab2: (A)** Significantly up-regulated host transcripts in the bacteriome tissue. **(B)** Significantly down-regulated host transcripts in the bacteriome tissue (compared to the rest of the gut). 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 four-fold expression change (logFC values shown) with p-value (Pvalue) 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.

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 to the haemolymph 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. An alternative explanation is that these cells are so tightly packed with endosymbionts that they are no longer able to function as regular midgut cells and are thus acting as a specialized space for harboring endosymbionts, with only a few essential host genes up-regulated. Since these two scenarios are not mutually exclusive, we recognize that the outcome is likely a gene/pathway-specific trade off between these two (and likely a few additional) explanations.

**Metal metabolism – comparison to tsetse flies and implications for blood digestion.** 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. 1999). *A. 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*).



**Fig3 (A): Schematic reconstruction of nutritional and immunity processes revealed as involved in host-symbiont interactions in the *M. 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. **(B): B-vitamin biosynthetic pathways and transporters in *A. melophagi*.** Genes are filled by heatmap colors representing level of expression (mean FPKM from 5 replicates).

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, 1998, McCall et al. 2000). Several transcripts for zinc transporters are highly up-regulated in bacteriocytes and very lowly expressed in the gut (Tab 2). Furthermore, the zinc ABC transporter of *A. melophagi* symbionts is highly

expressed (Fig 3AB; Suppl. 1) 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 (Tab 2; Suppl. 2), suggesting that zinc is mostly used by symbionts in the bacteriome section of the midgut.

A question of particular importance is which *A. melophagi* enzymes are responsible for this zinc dependence on their host and how symbionts participate on the blood digestion. There are numerous zinc-dependent enzyme candidates obvious from our data (Suppl. 1), particularly several zinc proteases (TldD, TldE, HtpX, FtsH, RseP, YebA) and a putative metallo-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. Metallo-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). Since 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 metallo-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.

**Insect immunity, peritrophic matrix, unknown proteins, and the role of gene specialization.** Among the transcripts up-regulated in the whole gut and down-regulated in bacteriocytes (Suppl. 2), 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 3A). However, their exact role (if any) in the control of *S. melophagi* and *B. melophagi* remains to be elucidated, although both can be highly effective on environmental bacteria. Strikingly, PGRP-LB, although highly expressed in both tissues, is not up-regulated in bacteriocytes (Fig 3A, Suppl. 3) 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, since 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. The latter explanation would be supported by the lack of peritrophic 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 do not include chitin synthesis genes of insect origin expressed in the gut, corroborating the microscopy observations.

As usual in insect RNA-Seq studies, many differentially expressed and/or highly expressed genes observed in our data (Suppl. 2) code conserved proteins with unknown function, hypothetical proteins, or orphan proteins with no or too low homology to proteins in protein databases. At least some of these genes are very likely involved in symbiosis maintenance and need to be examined experimentally in future studies. For example, RNA-Seq study on aphid bacteriocytes and whole mount *in situ* hybridizations of over-represented transcripts encoding aphid-specific orphan proteins has revealed a novel family of small cysteine-rich proteins with signal peptides (Shigenobu and Stern 2013). This finding supports hypotheses that numerous orphan genes in the pea aphid genome (International Aphid Genomics Consortium 2010) evolved to assist in lineage specific traits, such as symbiosis. A very similar scenario can be possibly applied to other symbiotic insects. The role of gene or even isoform specialization and recruitment for functioning in bacteriocytes was observed also in our data, for example in zinc transporters and proteinases (Suppl. 2).

**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 biosynthesis of EAAs by providing the starting products and carrying out the last aminotransferase steps of the pathways, we found no evidence for such intimate interactions in our *M. ovinus* data for the B-vitamin or other biosynthetic pathways (Tab 1; Fig 3; Suppl. 2). 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 hosts), 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. 2013, Duncan et al. 2014). This situation then inevitably leads to highly interconnected relationships. Moreover, if the bacteriome section is localized in the gut (e.g. in tsetse flies, louse flies, and carpenter ants), it cannot be fully specialized for interactions with the symbionts, because it still has to express at least a handful of genes for digestion of its food source. Evaluating the exact impact of this less intimate symbiotic integration can eventually help to untangle the gradual steps in genome reduction observed for endosymbiont genomes.

## Conclusions

Using RNA-Seq with five biological replicates of the midgut cells housing bacteria and the rest of the gut, we uncovered the interactions among a blood-sucking parasite and its microbiome. 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 this blood-sucking insect is much less intimately involved in cooperation on biosynthesis of nutrients than plant-sap sucking insects such as aphids, psyllids, whiteflies, and mealybugs. 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. Since 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.

## 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 DNase-d 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 (2x100 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 and generate a sufficient gene set. 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 [<http://code.google.com/p/cutadapt/>]. 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>] with a minimum window quality score of 20. Reads shorter than 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 *Arsenophonus 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 was subsequently improved according to the expression data, i.e. 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>]. Identification of transcript boundaries, quantification of transcript abundance (FPKM values), and prediction of operon structure was carried out in Rockhopper v2.0.3 (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 mapped 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. Only transcripts with at least four-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 (*Trypanosoma melophagium*) were removed from the reference transcriptome. However, we note that some of *Trypanosoma melophagium* transcripts could be missed by this approach since transcripts with no hits were retained. Single-gene phylogenetic inference would be needed to fully decontaminate *T. melophagium* contamination from the *Melophagus* reference transcriptome since no reference genome exists for neither *M. ovinus* nor *T. melophagium*. Since no high-quality genome data are available for *M. ovinus*, *T. melophagium*, *S. melophagi*, and *B. melophagi*, we did not inspect further the bin of all transcripts filtered out as bacteria (e.g. to detect possible horizontal gene transfers).

### *Transcriptome annotation and protein prediction*

TransDecoder [<https://github.com/TransDecoder/>]

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 B-vitamin pathways were updated with this expression information. Expression of genes annotated in the *A. melophagi* genome as transporters was assessed to reconstruct 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 co-factor 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 non-redundant 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/>) were used as queries. Pathways involved in control of symbiotic bacteria in gut and bacteriome tissues were compared to the situation in tsetse flies using literature review.

### **Description of additional data files:**

**Suppl. 1** Table with symbiont FPKM values for individual replicates, mean FPKM values, COG categories, locus tags, gene names, and protein products for the *A. melophagi* genome.

**Suppl. 2** Table with raw counts of mapped reads, normalized FPKM values, and EdgeR results for *M. ovinus*.

**Suppl. 3** Figure and table with IMD, JAK/STAT, JNK, and Toll genes detected in the *M. ovinus* transcriptome.

Additional large supplementary files are available on FigShare [ <https://doi.org/10.6084/m9.figshare.6146777.v1>]. In particular, the improved *A. melophagi* genome annotation and pathway maps with highlighted expression, *M. ovinus* transcriptome annotation report, total meta-transcriptome assembly (including predicted proteins), filtered reference transcriptome (including predicted proteins), complete results of the Trinity-Transdecoder-RSEM-edgeR-Trinotate pipeline (as a TrinotateWeb MySQL database), and supplementary figures with quality check analyses for individual samples/replicates and differential expression visualization (Volcano and MA plots).

### **Author's contributions:**

FH designed and coordinated the study, carried out the preparation of samples, analyses of data, and drafted the manuscript. VH carried out analyses of immune genes and participated in the study design and manuscript revisions. AD designed the study and participated in manuscript revisions. All authors read and approved the final manuscript.

### **Competing interest:**

The authors declare that they have no competing interests.

### **Acknowledgements:**

We thank the Centre for Genomic Research (University of Liverpool) for sequencing services. FH was supported by a postdoctoral research fellowship from EMBO (ALTF 1260-2016) while writing this article.

### **Funding:**

VH was supported by the Czech Science Foundation grant 18-07711S.

### **List of abbreviations used:**

COG: Clusters of Orthologous Groups  
 EAAs: Essential Amino Acids  
 FPKM: Fragments Per Kilobase Million  
 GGBP: Gram-Negative Binding Protein  
 HGT: Horizontal Gene Transfer  
 TPM: Transcripts Per Kilobase Million  
 TMM: Trimmed Mean of M-values normalization method  
 NR and NT: Non-redundant protein and nucleotide databases  
 ORF: Open Reading Frame  
 RNA-Seq: RNA Sequencing  
 PGRP: Peptidoglycan Recognition Protein  
 IMD: Immunodeficiency signaling pathway

## References

- Aksoy, S. 1995. Molecular analysis of the endosymbionts of tsetse flies: 16S rDNA locus and over-expression of a chaperonin. *Insect Molecular Biology* 4:23–29.
- Anselme, C., V. Perez-Brocal, A. Vallier, C. Vincent-Monegat, D. Charif, A. Latorre, A. Moya, and A. Heddi. 2008. Identification of the Weevil immune genes and their expression in the bacteriome tissue. *BMC Biology* 6:43.
- Anselme, C., A. Vallier, S. Balmand, M. O. Fauvarque, and A. Heddi. 2006. Host PGRP gene expression and bacterial release in endosymbiosis of the weevil *Sitophilus zeamais*. *Applied and Environmental Microbiology* 72:6766–6772.
- Baumann, P., L. Baumann, and M. A. Clark. 1996. Levels of *Buchnera aphidicola* chaperonin GroEL during growth of the aphid *Schizaphis graminum*. *Current Microbiology* 32:279–285.
- Belda, E., F. J. Silva, J. Pereto, and A. Moya. 2012. Metabolic networks of *Sodalis glossinidius*: a systems biology approach to reductive evolution. *Plos One* 7:e30652.
- Bing, X., G. M. Attardo, A. Vigneron, E. Aksoy, F. Scolari, A. Malacrida, B. L. Weiss, and S. Aksoy. 2017. Unravelling the relationship between the tsetse fly and its obligate symbiont *Wigglesworthia*: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proceedings of the Royal Society B-Biological Sciences* 284:20170360.
- Boyd, B. M., J. M. Allen, R. Koga, T. Fukatsu, A. D. Sweet, K. P. Johnson, and D. L. Reed. 2016. Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Applied and Environmental Microbiology* 82:3185–3197.
- Bradford, P. A. 2001. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clinical Microbiology Reviews* 14:933–951.
- Brucker, R. M., L. J. Funkhouser, S. Setia, R. Pauly, and S. R. Bordenstein. 2012. Insect Innate Immunity Database (IIID): an annotation tool for identifying immune genes in insect genomes. *Plos One* 7:e45125.
- Carver, T., S. R. Harris, M. Berriman, J. Parkhill, and J. A. McQuillan. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464–469.
- Charles, H., S. Balmand, A. Lamelas, L. Cottret, V. Perez-Brocal, B. Burdin, A. Latorre, G. Febvay, S. Colella, F. Calevro, and Y. Rahbe. 2011. A genomic reappraisal of symbiotic function in the aphid/*Buchnera* symbiosis: reduced transporter sets and variable membrane organisations. *Plos One* 6:e29096.
- Chrudimský, T., F. Husník, E. Nováková, and V. Hypša. 2012. *Candidatus Sodalis melophagi* sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. *Plos One* 7:e40354.
- Coleman, J. E. 1992. Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annual Review of Biochemistry* 61:897–946.
- Coleman, J. E. 1998. Zinc enzymes. *Current Opinion in Chemical Biology* 2:222–234.
- Crawford, J. M., R. Kontnik, and J. Clardy. 2010. Regulating alternative lifestyles in entomopathogenic bacteria. *Current biology : CB* 20:69–74.
- Douglas, A. E. 2016. How multi-partner endosymbioses function. *Nature Reviews Microbiology* 14:731–743.
- Drawz, S. M., and R. A. Bonomo. 2010. Three decades of beta-lactamase inhibitors. *Clinical Microbiology Reviews* 23:160–201.
- Duncan, R. P., F. Husnik, J. T. Van Leuven, D. G. Gilbert, L. M. Dávalos, J. P. McCutcheon, and A. C. C. Wilson. 2014. Dynamic recruitment of amino acid transporters to the insect/symbiont interface. *Molecular Ecology* 23:1608–23.
- Eichler, S., and G. A. Schaub. 2002. Development of symbionts in triatomine bugs and the effects of infections with trypanosomatids. *Experimental parasitology* 100:17–27.
- Fan, Y., J. W. Thompson, L. G. Dubois, M. A. Moseley, and J. J. Wernegreen. 2013. Proteomic analysis of an unculturable bacterial endosymbiont (*Blochmannia*) reveals high abundance of chaperonins and biosynthetic enzymes. *Journal of Proteome Research* 12:704–718.
- Fischer, J. D., G. L. Holliday, and J. M. Thornton. 2010. The CoFactor database: organic cofactors in enzyme catalysis. *Bioinformatics* 26:2496–2497.
- Fukatsu, T., and H. Ishikawa. 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. *Journal of Molecular Evolution* 36:568–577.
- Gerardo, N. M., B. Altincicek, C. Anselme, H. Atamian, S. M. Barribeau, M. de Vos, E. J. Duncan, J. D. Evans, T. Gabaldon, M. Ghanim, A. Heddi, I. Kaloshian, A. Latorre, A. Moya, A. Nakabachi, B. J. Parker, V. Perez-Brocal, M. Pignatelli, Y. Rahbe, J. S. Ramsey, C. J. Spragg, J. Tamames, D. Tamarit, C. Tamborindeguy, C. Vincent-Monegat, and A. Vilcinskis. 2010. Immunity and other defenses in pea aphids, *Acyrtosiphon pisum*. *Genome Biology* 11:R21.
- Gerardo, N. M., and A. C. Wilson. 2011. The power of paired genomes. *Molecular Ecology* 20:2038–2040.
- Gonzalez-Domenech, C. M., E. Belda, R. Patino-Navarrete, A. Moya, J. Pereto, and A. Latorre. 2012. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two *Blattabacterium cuenoti* strains, primary endosymbionts of cockroaches. *BMC Microbiology* 12 Suppl 1:S5.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* 29:644–652.
- Gruber-Vodicka, H. R., B. K. Seah, and E. Pruesse. 2019. phyloFlash — Rapid SSU rRNA profiling and targeted assembly from metagenomes. *bioRxiv*:521922.
- Hansen, A. K., and N. A. Moran. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences of the United States of America* 108:2849–2854.
- Hansen, A. K., and N. A. Moran. 2013. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular Ecology* 23:1473–96.

- Husnik, F. 2018. Host-symbiont-pathogen interactions in blood-feeding parasites: nutrition, immune cross-talk and gene exchange. *Parasitology* 145:1294–1303.
- Husnik, F., and J. P. McCutcheon. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* 113:E5416–E5424.
- Husnik, F., N. Nikoh, R. Koga, L. Ross, R. P. Duncan, M. Fujie, M. Tanaka, N. Satoh, D. Bachtrog, A. C. C. Wilson, C. D. von Dohlen, T. Fukatsu, and J. P. McCutcheon. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153:1567–1578.
- International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biology* 8:e1000313.
- Karp, P. D., S. M. Paley, M. Krummenacker, M. Latendresse, J. M. Dale, T. J. Lee, P. Kaipa, F. Gilham, A. Spaulding, L. Popescu, T. Altman, I. Paulsen, I. M. Keseler, and R. Caspi. 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* 11:40–79.
- Kirkness, E. F., B. J. Haas, W. L. Sun, H. R. Braig, M. A. Perotti, J. M. Clark, S. H. Lee, H. M. Robertson, R. C. Kennedy, E. Elhaik, D. Gerlach, E. V. Kriventseva, C. G. Elsik, D. Graur, C. A. Hill, J. A. Veenstra, B. Walenz, J. M. C. Tubio, J. M. C. Ribeiro, J. Rozas, J. S. Johnston, J. T. Reese, A. Popadic, M. Tojo, D. Raoult, D. L. Reed, Y. Tomoyasu, E. Krause, O. Mittapalli, V. M. Margam, H. M. Li, J. M. Meyer, R. M. Johnson, J. Romero-Severson, J. P. VanZee, D. Alvarez-Ponce, F. G. Vieira, M. Aguade, S. Guirao-Rico, J. M. Anzola, K. S. Yoon, J. P. Strycharz, M. F. Unger, S. Christley, N. F. Lobo, M. J. Seufferheld, N. K. Wang, G. A. Dasch, C. J. Struchiner, G. Madey, L. I. Hannick, S. Bidwell, V. Joardar, E. Caler, R. F. Shao, S. C. Barker, S. Cameron, R. V. Bruggner, A. Regier, J. Johnson, L. Viswanathan, T. R. Utterback, G. G. Sutton, D. Lawson, R. M. Waterhouse, J. C. Venter, R. L. Strausberg, M. R. Berenbaum, F. H. Collins, E. M. Zdobnov, and B. R. Pittendrigh. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America* 107:12168–12173.
- Laetsch, D. R., and M. L. Blaxter. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research* 6:1287.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
- Lehane, M. J. 1997. Peritrophic matrix structure and function. *Annual Review of Entomology* 42:525–550.
- Van Leuven, J. T., R. C. Meister, C. Simon, and J. P. McCutcheon. 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158:1270–1280.
- Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
- Login, F. H., S. Balmand, A. Vallier, C. Vincent-Monegat, A. Vigneron, M. Weiss-Gayet, D. Rochat, and A. Heddi. 2011. Antimicrobial peptides keep insect endosymbionts under control. *Science* 334:362–365.
- Lu, H. ling, C. che Chang, and A. C. C. Wilson. 2016. Amino acid transporters implicated in endocytosis of *Buchnera* during symbiont transmission in the pea aphid. *EvoDevo* 7:24.
- Luan, J.-B., W. Chen, D. K. Hasegawa, A. M. Simmons, W. M. Wintermantel, K.-S. Ling, Z. Fei, S.-S. Liu, and A. E. Douglas. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biology and Evolution* 7:2635–47.
- Macdonald, S. J., G. G. Lin, C. W. Russell, G. H. Thomas, and A. E. Douglas. 2012. The central role of the host cell in symbiotic nitrogen metabolism. *Proceedings of the Royal Society B-Biological Sciences* 279:2965–2973.
- MacDonald, S. J., G. H. Thomas, and A. E. Douglas. 2011. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Molecular Ecology* 20:2073–2084.
- Mantilla, B. S., L. Marchese, A. Casas-Sánchez, N. A. Dyer, N. Ejeh, M. Biran, F. Bringaud, M. J. Lehane, A. Acosta-Serrano, and A. M. Silber. 2017. Proline metabolism is essential for *Trypanosoma brucei brucei* survival in the tsetse vector. *PLoS Pathogens* 13:1–29.
- Mao, M., X. Yang, and G. M. Bennett. 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proceedings of the National Academy of Sciences of the United States of America* 115:E11691–E11700.
- McCall, K. A., C. Huang, and C. A. Fierke. 2000. Function and mechanism of zinc metalloenzymes. *Journal of Nutrition* 130:1437S–46S.
- McClure, R., D. Balasubramanian, Y. Sun, M. Bobrovskyy, P. Sumby, C. A. Genco, C. K. Vanderpool, and B. Tjaden. 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research* 41:e140.
- McCutcheon, J., B. McDonald, and N. Moran. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proceedings of the National Academy of Sciences of the United States of America* 106:15394–15399.
- McCutcheon, J. P., and N. A. Moran. 2011. Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* 10:13–26.
- Michalkova, V., J. B. Benoit, B. L. Weiss, G. M. Attardo, and S. Aksoy. 2014. Obligate symbiont-generated vitamin B6 is critical to maintain proline homeostasis and fecundity in tsetse flies. *Applied and environmental microbiology:AEM.01150-14-*.
- Moran, N. A., and G. M. Bennett. 2014. The tiniest tiny genomes. *Annual Review of Microbiology* 68:195–215.
- Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* 42:165–190.
- Nakabachi, A., K. Ishida, Y. Hongoh, M. Ohkuma, and S. Miyagishima. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Current Biology* 24:R640-641.
- Nakabachi, A., and H. Ishikawa. 1999. Provision of riboflavin to the host aphid, *Acyrtosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *Journal of Insect Physiology* 45:1–6.
- Nakabachi, A., S. Shigenobu, N. Sakazume, T. Shiraki, Y. Hayashizaki, P. Carninci, H. Ishikawa, T. Kudo, and T. Fukatsu. 2005. Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proceedings of the National Academy of Sciences of the United States of America* 102:5477–5482.

- Nikoh, N., T. Hosokawa, M. Moriyama, K. Oshima, M. Hattori, and T. Fukatsu. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences* 111:1409284111.
- Nováková, E., F. Husník, E. Šochová, and V. Hypša. 2015. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and Environmental Microbiology* 81:6189–6199.
- Nováková, E., V. Hypša, P. Nguyen, F. Husník, and A. C. Darby. 2016. Genome sequence of *Candidatus Arsenophonus lipopteni*, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *Standards in Genomic Sciences* 11.
- Oakeson, K. F., R. Gil, A. L. Clayton, D. M. Dunn, A. C. von Niederhausern, C. Hamil, A. Aoyagi, B. Duval, A. Baca, F. J. Silva, A. Vallier, D. G. Jackson, A. Latorre, R. B. Weiss, A. Heddi, A. Moya, and C. Dale. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biology and Evolution* 6:76–93.
- Pais, R., C. Lohs, Y. Wu, J. Wang, and S. Aksoy. 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology* 74:5965–5974.
- Poliakov, A., C. W. Russell, L. Ponnala, H. J. Hoops, Q. Sun, A. E. Douglas, and K. J. van Wijk. 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. *Molecular & Cellular Proteomics* 10:M110.007039.
- Price, D. R. G., H. Feng, J. D. Baker, S. Bavan, C. W. Luetje, and A. C. C. Wilson. 2013. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences of the United States of America* 111:320–5.
- Ratzka, C., R. Gross, and H. Feldhaar. 2013. Gene expression analysis of the endosymbiont-bearing midgut tissue during ontogeny of the carpenter ant *Camponotus floridanus*. *Journal of Insect Physiology* 59:611–623.
- Říhová, J., E. Nováková, F. Husník, and V. Hypša. 2017. *Legionella* becoming a mutualist: Adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome Biology and Evolution* 9:2946–2957.
- Rio, R. V., R. E. Symula, J. Wang, C. Lohs, Y. N. Wu, A. K. Snyder, R. D. Bjornson, K. Oshima, B. S. Biehl, N. T. Perna, M. Hattori, and S. Aksoy. 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *MBio* 3:1–13.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Roje, S. 2007. Vitamin B biosynthesis in plants. *Phytochemistry* 68:1904–1921.
- Shigenobu, S., and D. L. Stern. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proc Biol Sci* 280:20121952.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)* 31:3210–3212.
- Sloan, D. B., A. Nakabachi, S. Richards, J. Qu, S. C. Murali, R. A. Gibbs, and N. A. Moran. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution* 31:857–871.
- Small, R. W. 2005. A review of *Melophagus ovinus* (L.), the sheep ked. *Veterinary Parasitology* 130:141–155.
- Soulages, E. L. A. and J. L. 2010. Insect fat body: Energy, metabolism, and regulation. *Annual Review of Entomology* 55:207–225.
- Strickler-Dinglasan, P. M., N. Guz, G. Attardo, and S. Aksoy. 2006. Molecular characterization of iron binding proteins from *Glossina morsitans morsitans* (Diptera: Glossinidae). *Insect Biochemistry and Molecular Biology* 36:921–933.
- Thomas, G. H., J. Zucker, S. J. Macdonald, A. Sorokin, I. Goryanin, and A. E. Douglas. 2009. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Systematic Biology* 3:24.
- Vigneron, A., F. Masson, A. Vallier, S. Balmand, M. Rey, C. Vincent-Monégat, E. Aksoy, E. Aubailly-Giraud, A. Zaidman-Rémy, and A. Heddi. 2014. Insects recycle endosymbionts when the benefit is over. *Current biology : CB* 24:2267–73.
- Wang, J. W., and S. Aksoy. 2012. PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. *Proceedings of the National Academy of Sciences* 109:10552–10557.
- Wang, J. W., Y. N. Wu, G. X. Yang, and S. Aksoy. 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proceedings of the National Academy of Sciences of the United States of America* 106:12133–12138.
- Waterhouse, D. F. 1953. The occurrence and significance of the peritrophic membrane, with special reference to adult lepidoptera and diptera. *Australian Journal of Zoology* 1:299–318.
- Weiss, B. L., J. Wang, and S. Aksoy. 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biology* 9:e1000619.
- Yan, J., Q. Cheng, C. B. Li, and S. Aksoy. 2002. Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. *Insect Molecular Biology* 11:57–65.
- Yoshiga, T., T. Georgieva, B. C. Dunkov, N. Harizanova, K. Ralchev, and J. H. Law. 1999. *Drosophila melanogaster* transferrin - Cloning, deduced protein sequence, expression during the life cycle, gene localization and up-regulation on bacterial infection. *European Journal of Biochemistry* 260:414–420.