RESEARCH ARTICLE

The holobiont transcriptome of teneral tsetse fly species of varying vector competence

Miguel Medina Munoz[1], Caitlyn Brenner[2], Dylan Richmond[1], Noah Spencer[1] and Rita V. M. Rio[1*]

Abstract

Background: Tsetse flies are the obligate vectors of African trypanosomes, which cause Human and Animal African Trypanosomiasis. Teneral flies (newly eclosed adults) are especially susceptible to parasite establishment and development, yet our understanding of why remains fragmentary. The tsetse gut microbiome is dominated by two Gammaproteobacteria, an essential and ancient mutualist *Wigglesworthia glossinidia* and a commensal *Sodalis glossinidius*. Here, we characterize and compare the metatranscriptome of teneral *Glossina morsitans* to that of *G. brevipalpis* and describe unique immunological, physiological, and metabolic landscapes that may impact vector competence differences between these two species.

Results: An active expression profile was observed for *Wigglesworthia* immediately following host adult metamorphosis. Specifically, 'translation, ribosomal structure and biogenesis' followed by 'coenzyme transport and metabolism' were the most enriched clusters of orthologous genes (COGs), highlighting the importance of nutrient transport and metabolism even following host species diversification. Despite the significantly smaller *Wigglesworthia* genome more differentially expressed genes (DEGs) were identified between interspecific isolates ($n = 326$, ~ 55% of protein coding genes) than between the corresponding *Sodalis* isolates ($n = 235$, ~ 5% of protein coding genes) likely reflecting distinctions in host co-evolution and adaptation. DEGs between *Sodalis* isolates included genes involved in chitin degradation that may contribute towards trypanosome susceptibility by compromising the immunological protection provided by the peritrophic matrix. Lastly, *G. brevipalpis* tenerals demonstrate a more immunologically robust background with significant upregulation of IMD and melanization pathways.

Conclusions: These transcriptomic differences may collectively contribute to vector competence differences between tsetse species and offers translational relevance towards the design of novel vector control strategies.

Keywords: Symbiosis, RNA-Seq, *Wigglesworthia*, *Sodalis*, Tsetse, Teneral, *Glossina*, Vector competence

Background

Tsetse flies (Diptera: Glossinidae) inhabit much of sub-Saharan Africa in an area referred to as the "tsetse belt", where significant detriment towards public health and agricultural sustainability occur due to the presence of these vectors [1, 2]. Tsetse adults of both sexes are strictly hematophagous, and thus, have epidemiological significance as the cyclical (and obligate) vectors of human and animal African trypanosomes. Because there are no vaccines and few pharmaceuticals available, vector control [3–5] remains a significant component of programs intended to impede the transmission of trypanosome infections.

* Correspondence: rita.rio@mail.wvu.edu
[1]Department of Biology, Eberly College of Arts and Sciences, West Virginia University, Morgantown, WV 26505, USA
Full list of author information is available at the end of the article

Medina Munoz *et al. BMC Genomics* (2021) 22:400

Page 2 of 20

Tsetse flies have a viviparous reproductive biology [6] characterized by the 'in utero' development of a single larva during each gonotrophic cycle. Here, the larva receives nutritional lipids and proteins through maternal secretions from modified female accessory reproductive glands known as milk glands. These milk gland secretions also seed progeny with the core bacteria of the tsetse digestive tract microbiota [7–9], specifically the obligate mutualist *Wigglesworthia glossinidia* [10] and the commensal *Sodalis glossinidius* [11] (hereafter *Wigglesworthia* and *Sodalis*, respectively). Although a more complex microbial diversity in the digestive tracts of adult flies has been reported, these environmentally acquired bacteria are lacking within tenerals (newly emerged adults which have not yet fed), are transient, and not integrated into tsetse biology due to their random occurrence [12–14].

Despite sharing a common route of maternal inheritance, the *Wigglesworthia* and *Sodalis* endosymbionts have distinct coevolutionary histories with their tsetse host. The [14] ancient mutualist, *Wigglesworthia*, likely established at the commencement (or soon thereafter) of tsetse species radiation with subsequent codiversification across host species spanning the course of 50–80 million years [15]. The extraordinary significance of the *Wigglesworthia* mutualism towards tsetse biology is symbolized by the bacteriome structure [16], which exclusively harbors *Wigglesworthia* intracellularly within specialized tsetse epithelial cells (known as bacteriocytes) in an immunotolerant niche [17] in return for nutrient supplementation of the limited blood diet. In contrast, *Sodalis* symbionts have a wide tropism being found in the tsetse digestive tract, muscle, hemolymph, salivary glands and fat body [9, 18]. Unlike *Wigglesworthia*, *Sodalis* has established much more recently as reflected by its stochastic distribution in wild tsetse populations [19–23] which also indicates that this symbiosis is not a requisite for tsetse survival although its role towards tsetse vector competence remains contentious [24].

Age impacts tsetse vector competence as teneral flies have higher permissiveness towards trypanosome infections upon acquisition of their first blood meal (referred to as the "teneral phenomenon" reviewed in [25]). Although numerous microbiota contributions towards tsetse biology have been functionally characterized [26–28], the contributions of *Wigglesworthia* and *Sodalis* to the "teneral phenomenon" remain poorly understood. Here, we characterize and compare the *Wigglesworthia*, *Sodalis* and tsetse transcriptomes within teneral flies of low (*Glossina brevipalpis*) versus high (*G. morsitans*) vector competence. We present differences in the gene expression profiles of tsetse, *Wigglesworthia* and *Sodalis* that likely contribute towards unique immunological, physiological and metabolic landscapes that may impact vector competence. In particular, we focus on genes involved in nutrient provisioning with *Wigglesworthia*, genes that may disrupt tsetse physiology and facilitate trypanosome infections with *Sodalis*, and immunity in teneral tsetse of different species. These findings have significance not only towards understanding the basic biology of tsetse and its microbiota during the incipient stages of adulthood, but also applied merit towards identifying biological factors which may predispose certain tsetse species towards higher trypanosome infection rates which can be used in the design of novel control mechanisms.

Results

General transcriptome features

Eighteen cDNA libraries were generated from the total homogenates of either sex-specific bacteriomes or midguts from *G. brevipalpis* and *G. morsitans* teneral (1 day old, unfed virgin) tsetse flies. For each species, we had a collection of 6 bacteriome libraries and 3 midgut libraries. Each library consisted of a pool of either 20 bacteriomes or midguts. The libraries were sequenced using Illumina HiSeq technology, resulting in an average of 17,390,200 ± 2,558,126 (Std. dev) of paired-end reads (2 × 51 bp in length) from each sample. All reads were scored as "very good" quality (Phred scores > 30, Additional file 1: Figure S1). Reads were mapped to a reference dataset consisting of the collective genomes of tsetse species (*G. brevipalpis* and *G. morsitans*), *Wigglesworthia* isolates (*W. glossinidia morsitans* and *W. glossinidia brevipalpis*), and *S. glossinidius* (Additional file 1: Figure S2) using Salmon [29]. Importantly, the majority of reads mapped back only to their corresponding genomes serving as a control for specificity. A total of 112,895,620 reads (corresponding to ~ 63.2% of mapped reads) were identified as having tsetse origin, 65,550,668 reads (corresponding to ~ 36.7% of mapped reads) mapped to *Wigglesworthia* and 214,292 reads (corresponding to ~ 0.1% of mapped reads) were recognized as *Sodalis*. The lower abundance of *Sodalis* reads relative to *Wigglesworthia* likely arises due to a significantly lower population density at the teneral stage [30]. Nonspecific reads (i.e. reads that mapped back to the genomes which were not of origin) were excluded from further analyses. There were no significant differences in either the mean abundance of total reads (total *G. brevipalpis* mean reads $\pm$ SEM = 17,744,213 $\pm$ 889,174, total *G. morsitans* mean reads $\pm$ SEM = 17,036,187 $\pm$ 850,383, Unpaired Student's *t*-test, p = 0.57, Additional file 1: Figure S3A) or mapped reads (mapped *G. brevipalpis* mean reads $\pm$ SEM = 10,686,667 $\pm$ 823,288, mapped *G. morsitans* mean reads $\pm$ SEM = 9,164,444 $\pm$ 1,165,016, Mann-Whitney test, p = 0.42, Additional file 1: Figure S3B) between tsetse species libraries.

Wigglesworthia-based analyses

As *Wigglesworthia* shapes multiple aspects of tsetse physiology, immune system maturation and metabolism [17, 26–28, 31–36], we were particularly interested in identifying genes expressed by *Wigglesworthia* within teneral flies and determining whether these may differ between sexes and tsetse species. Our rationale for these analyses is that distinct expression profiles may likely be a product of host-symbiont interdependency, representing points of diversification in the symbiosis during the course of coevolution of tsetse and their respective *Wigglesworthia* which may then impact tsetse biology including vector competence.

Our initial comparison focused on highly expressed *Wigglesworthia* genes, which we defined as loci with expression levels of ≥ 100 Transcripts per million (TPM) (following [37–40]). Using these criteria 72–83% of the *Wigglesworthia* genome was highly expressed across all libraries indicating an active transcriptional profile for this symbiont upon tsetse adult metamorphosis (Fig. 1). Within species comparison demonstrated no significant difference in the mean *Wigglesworthia* expression between sexes for *G. morsitans* isolates (Wgm; female mean TPM $\pm$ SEM = 1639 $\pm$ 225.6, male mean TPM + SEM = 1562 + 252.1; Mann-Whitney test, p = 0.7865; Additional file 1: Fig. S4A). However, there was a slightly higher mean expression for *Wigglesworthia* genes within male bacteriomes of *G. brevipalpis* isolates (Wgb; mean female TPM + SEM = 1133 $\pm$ 46.01, mean male TPM $\pm$ SEM = 1239 $\pm$ 67.65; Mann-Whitney test, p = 0.022; Additional file 1: Fig. S4B). Upon comparison of bacteriome libraries between the tsetse species, mean *Wigglesworthia* expression is significantly higher within *G. morsitans* relative to *G. brevipalpis* libraries (Mean Wgm TPM $\pm$ SEM = 1613 $\pm$ 172.1, mean Wgb TPM $\pm$ SEM = 1169 $\pm$ 38.09); Mann-Whitney test, p < 0.0001; Additional file 1: Fig. S4C), supporting greater *Wigglesworthia* activity within *G. morsitans* and, although speculative, higher symbiont reliance.

Principal Component Analysis (PCA) was used to compare the global gene expression of *Wigglesworthia* between female and male libraries within a host species (Fig. 2A & B). In both tsetse species, the female and male *Wigglesworthia* libraries separated along the second principal component, which explained ~ 22% of variance in expression. An additional PCA on a core gene set (consisting of 591 genes) was performed to compare *Wigglesworthia* gene expression between the tsetse species. Upon the comparison of *Wigglesworthia* expression between *G. brevipalpis* and *G. morsitans* bacteriomes, species-specific libraries separated along the first principal component, which explained 71% of variance (Fig. 2C).

Tsetse sex drives differential expression of Wigglesworthia genes. DESeq [41] was used to identify differentially expressed genes (DEGs). A total of 26 DEGs were identified between *Wigglesworthia* libraries obtained from female versus male *G. morsitans* bacteriomes (Fig. 3A, Additional file 2: Table S1), with all of these significantly upregulated within female bacteriomes. Eleven of these genes (42%) are involved in metabolic roles including B vitamin synthesis such as *bioA* and *bioD,* both components of the biotin (B1) synthesis pathway, and *pdxB* in pyridoxal 5′-phosphate (B6) metabolism. The sigma 28 regulator of class III flagella genes (*fliA*) involved in controlling the assembly of the final components of flagellum and WIGMOR_RS00305, a homolog of *fliJ*, were also significantly increased in expression. The *fliJ* gene within *Wigglesworthia* isolated from *G. morsitans* was previously characterized as a pseudogene due to truncation [42], but a 41% amino acid identity between the two *Wigglesworthia* homologs with a particularly high retention in residue identity within the two critical binding domains (aa 39–51 for FlgN binding and aa 65–82 for FliT binding) [43] suggests at least some preservation of function which merits further investigation. The gene *purF*, also significantly upregulated, encodes amidophosphoribosyltransferase, the enzyme that catalyzes the initial step in de novo purine biosynthesis [44, 45].

An elevated expression of *Wigglesworthia hslU*, a homolog of a chaperone-related protease [46, 47], was also observed within *G. morsitans* female bacteriomes. Interestingly, the oligomerization of HslU subunits is associated with the regulation of cell growth [48] and may support the significant increase in *Wigglesworthia* density during early adulthood observed in females but lacking in male tsetse flies [30]. We further confirmed the upregulation of a subset of these *Wigglesworthia* genes (*fliA, purF, bioA* and *bioD*) within female bacteriomes through qRT-PCR (Fig. 3B). Although PCA analyses showed a sex separation between *G. brevipalpis* libraries, DESeq found no significant difference on a gene by gene basis when comparing *Wigglesworthia* expression.

Differential expression of Wigglesworthia *genes between tsetse species.* Within the 16 Clusters of Orthologous Groups (COG, [49]) that were shared between the two tsetse species, the proportion of genes within each COG did not significantly differ, neither for the highly expressed genes (TPM $\geq$ 100, Kolmogorov-Smirnov test, p > 0.9999, Fig. 4A), nor for the DEGs (Kolmogorov-Smirnov test, p = 0.9718; Fig. 4B). An active expression profile was observed for *Wigglesworthia* immediately following host adult metamorphosis. Specifically, 'translation, ribosomal structure and biogenesis' followed by 'coenzyme transport and metabolism' were the most enriched COGs, highlighting the preservation of
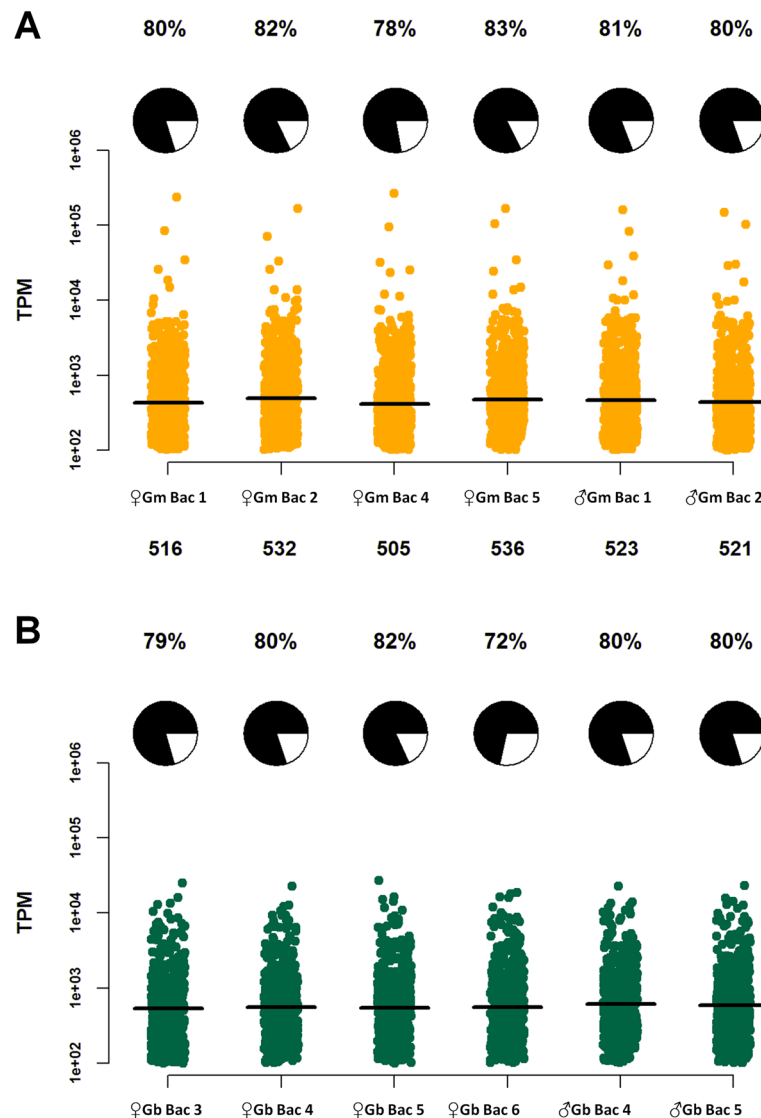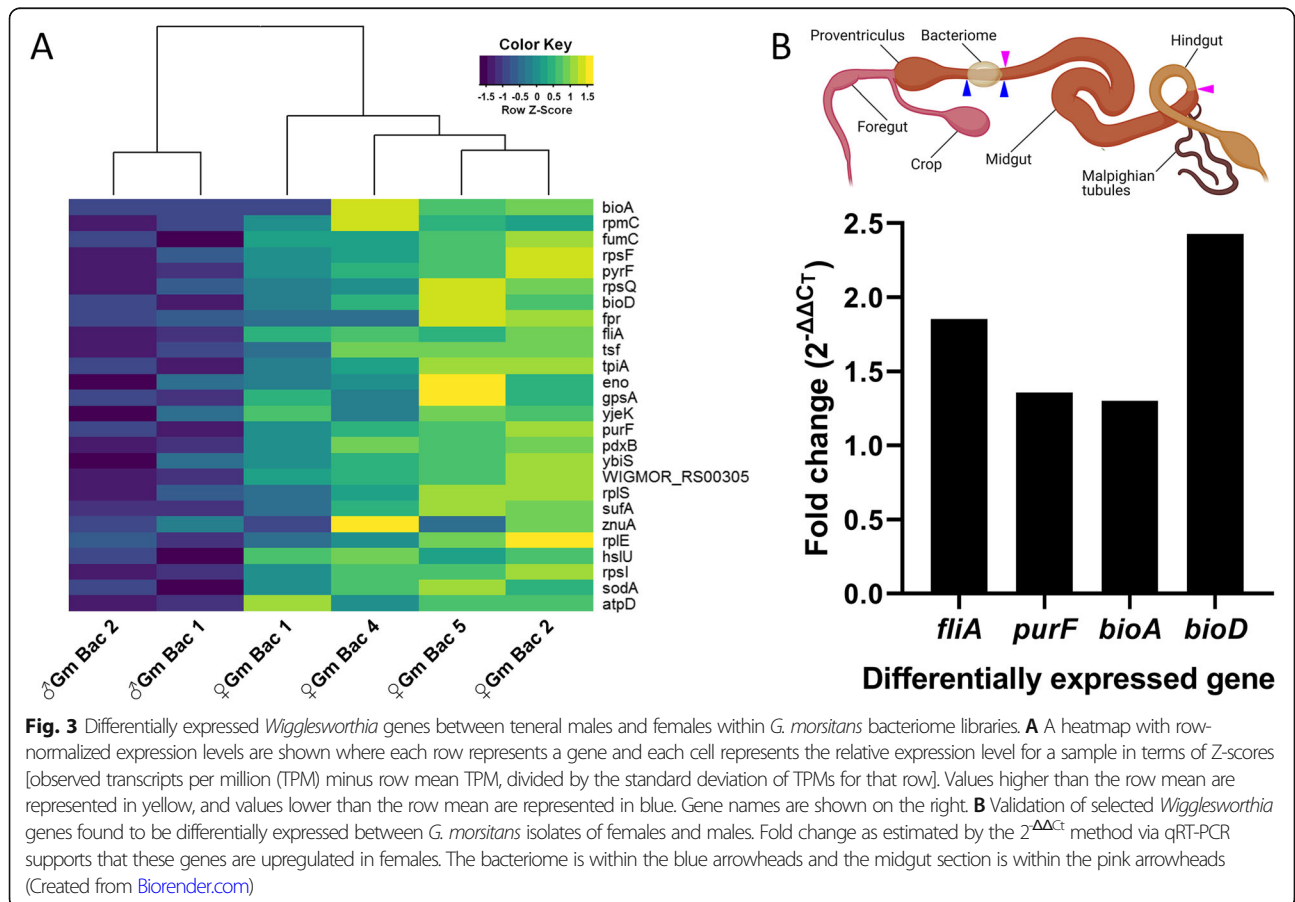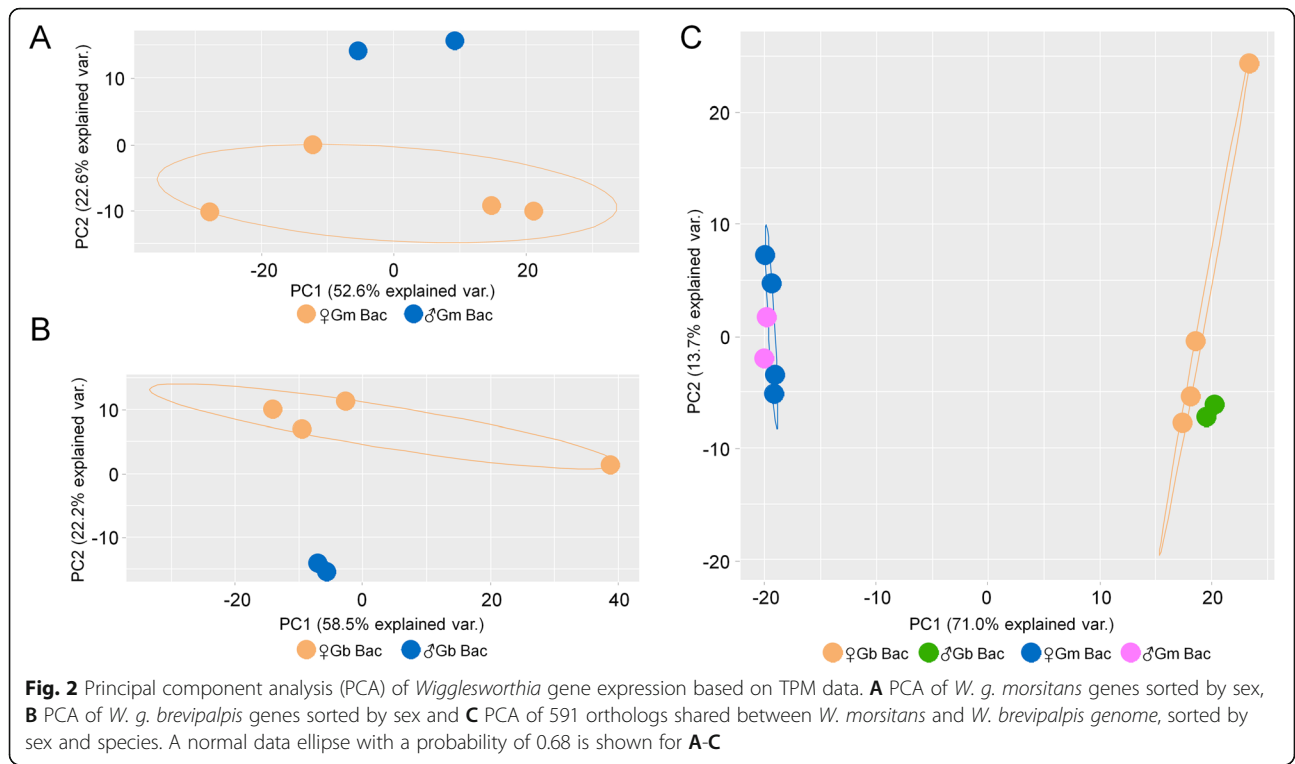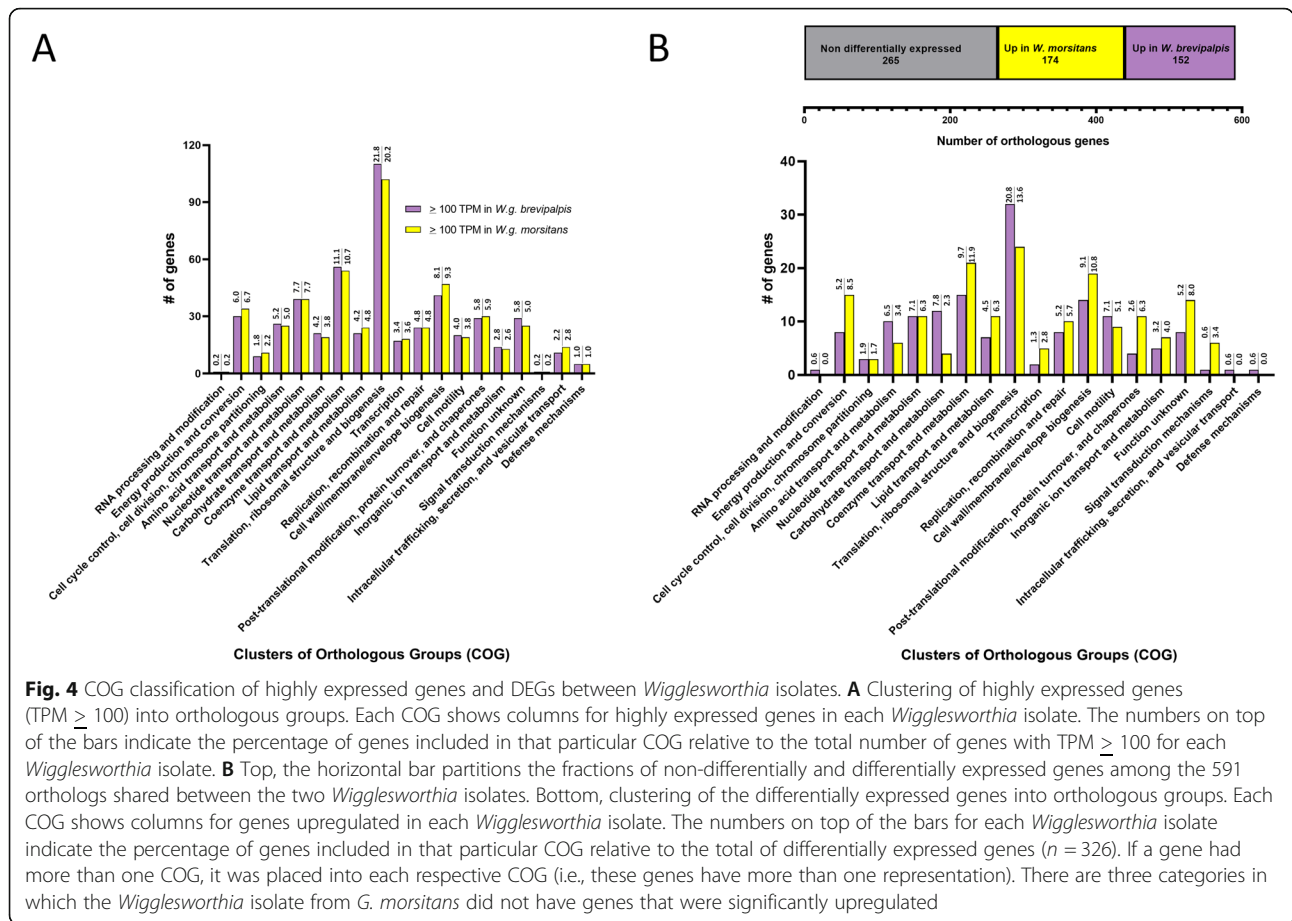
Fig. 1 *Wigglesworthia* gene expression per tsetse species. Median TPM values (for genes with TPM ≥ than 100) are shown as horizontal black lines for **A** *Wigglesworthia* within teneral *G. morsitans* bacteriomes and **B** *Wigglesworthia* within teneral *G. brevipalpis* bacteriomes. Dots around the median indicate TPM values of individual genes. The percentage of the total *Wigglesworthia* gene count represented by loci with TPM ≥ 100 is indicated as pie charts on top of each corresponding library. On the x axis, the identification of the library origin and the number of *Wigglesworthia* genes with TPM ≥ 100 is indicated. The y axis is in logarithmic scale

Wigglesworthia's vitamin provisioning role following host speciation. A total of 326 orthologs (55% of 591 protein coding genes, Fig. 4B) were identified as DEGs between the *Wigglesworthia* isolates of the two tsetse species (Additional file 3: Table S2). The *Wigglesworthia* symbionts within the *G. morsitans* bacteriomes demonstrate upregulation of 174 genes (53% of DEG genes, which corresponds to 29% of orthologous genes), while 152 genes (47% of DEG genes, which corresponds to 26% of orthologous genes) are upregulated within *G. brevipalpis* (Fig. 4B). Three COG categories which were unique to *G. brevipalpis* each only housed a single DEG

that was significantly upregulated; "RNA processing and modification" (*yjeR*), "Intracellular trafficking, secretion and vesicular transport" (*fliI*), and "Defense mechanisms" (*yadH*). Both *Wigglesworthia* isolates (Fig. 4B) had a significant proportion of differentially expressed genes (i.e. 22.1% of DEG) falling within COGs associated with "Energy production and conversion" and towards the "Transport and metabolism of ..." "lipids", "amino acids", and "carbohydrates". Additionally, ~ 11% of DEGs fell in the COG of "Coenzyme transport and metabolism".

The *Wigglesworthia* genomes encode a complete flagellar apparatus [42, 50] which likely facilitate its

Fig. 2 Principal component analysis (PCA) of *Wigglesworthia* gene expression based on TPM data. **A** PCA of *W. g. morsitans* genes sorted by sex, **B** PCA of *W. g. brevipalpis* genes sorted by sex and **C** PCA of 591 orthologs shared between *W. morsitans* and *W. brevipalpis* genome, sorted by sex and species. A normal data ellipse with a probability of 0.68 is shown for **A**-**C**

Fig. 3 Differentially expressed *Wigglesworthia* genes between teneral males and females within *G. morsitans* bacteriome libraries. **A** A heatmap with row-normalized expression levels are shown where each row represents a gene and each cell represents the relative expression level for a sample in terms of Z-scores [observed transcripts per million (TPM) minus row mean TPM, divided by the standard deviation of TPMs for that row]. Values higher than the row mean are represented in yellow, and values lower than the row mean are represented in blue. Gene names are shown on the right. **B** Validation of selected *Wigglesworthia* genes found to be differentially expressed between *G. morsitans* isolates of females and males. Fold change as estimated by the $2^{-\Delta\Delta Ct}$ method via qRT-PCR supports that these genes are upregulated in females. The bacteriome is within the blue arrowheads and the midgut section is within the pink arrowheads (Created from Biorender.com)

Fig. 4 COG classification of highly expressed genes and DEGs between *Wigglesworthia* isolates. **A** Clustering of highly expressed genes (TPM ≥ 100) into orthologous groups. Each COG shows columns for highly expressed genes in each *Wigglesworthia* isolate. The numbers on top of the bars indicate the percentage of genes included in that particular COG relative to the total number of genes with TPM ≥ 100 for each *Wigglesworthia* isolate. **B** Top, the horizontal bar partitions the fractions of non-differentially and differentially expressed genes among the 591 orthologs shared between the two *Wigglesworthia* isolates. Bottom, clustering of the differentially expressed genes into orthologous groups. Each COG shows columns for genes upregulated in each *Wigglesworthia* isolate. The numbers on top of the bars for each *Wigglesworthia* isolate indicate the percentage of genes included in that particular COG relative to the total of differentially expressed genes (*n* = 326). If a gene had more than one COG, it was placed into each respective COG (i.e., these genes have more than one representation). There are three categories in which the *Wigglesworthia* isolate from *G. morsitans* did not have genes that were significantly upregulated

evolutionary persistence through vertical transmission using a milk gland route [7–9]. Interestingly, the expression patterns of flagellar genes cluster by host species. A total of 22 genes out of 37 flagellum genes examined (~ 60%) significantly differed in their expression between the tsetse species, with one belonging to Class I, 18 to Class II and three to Class III (Fig. 5). The operon *flhDC* is a master regulator of flagellar genes [51] that activates the expression of Class II flagellar components. A significantly higher expression of the *Wigglesworthia flhDC* operon was observed in *G. morsitans* isolates. Counterintuitively, the corresponding expression levels of Class II flagellar genes downstream to *flhDC* were downregulated. Previous tsetse-*Wigglesworthia* studies demonstrated that the flagellar apparatus is downregulated within the bacteriome population, while it is actively synthesized by the milk gland population [42]. Interestingly, that previous study was based on the detection of the genes *fliC* and *motA*, which have a lower expression in our *G. morsitans* bacteriomes isolates, but a significantly higher expression in our *G. brevipalpis* isolates. A differential spatial and temporal regulation of flagella components of *Wigglesworthia* between tsetse host species merits further investigation.

Sodalis-based analyses

To further test the hypothesis that host species impact transcriptomic profiles in their microbiota and that these will be more pronounced the older the symbiosis, we also characterized and compared *Sodalis* gene expression in the two tsetse species. In all our midgut libraries, the reads mapping to the *Sodalis* genome constitute a very small proportion of total reads (~ 0.1%). The percentage of genes that show some level of expression varies widely across libraries, ranging from 58 to 64% in the *G. morsitans* midgut isolates, with an even greater span in the *G. brevipalpis* midgut isolates of 28–79% (over a total of 4541 protein coding genes) (Fig. 6A). If a gene was expressed, it likely had a transcription level of < 100 TPM (i.e. > 99% of 4541 genes, Fig. 6A). Only a small number of genes, constituting less than 1% of the total genes, is highly expressed (TPM ≥ 100). More specifically, the proportion of genes that were highly expressed is significantly larger in the *G. morsitans* isolates (26.33 ± 4.410) when compared to *G. brevipalpis* isolates (1.333 ± 0.333, Welch's test, *p* = 0.0291, Additional file 4: Table S3). For those genes with ≥ 100 TPM, the average TPM does not differ between the two tsetse host species (275.9 ± 26.87 in *G. morsitans* isolates vs. 208.2 ± 9.799
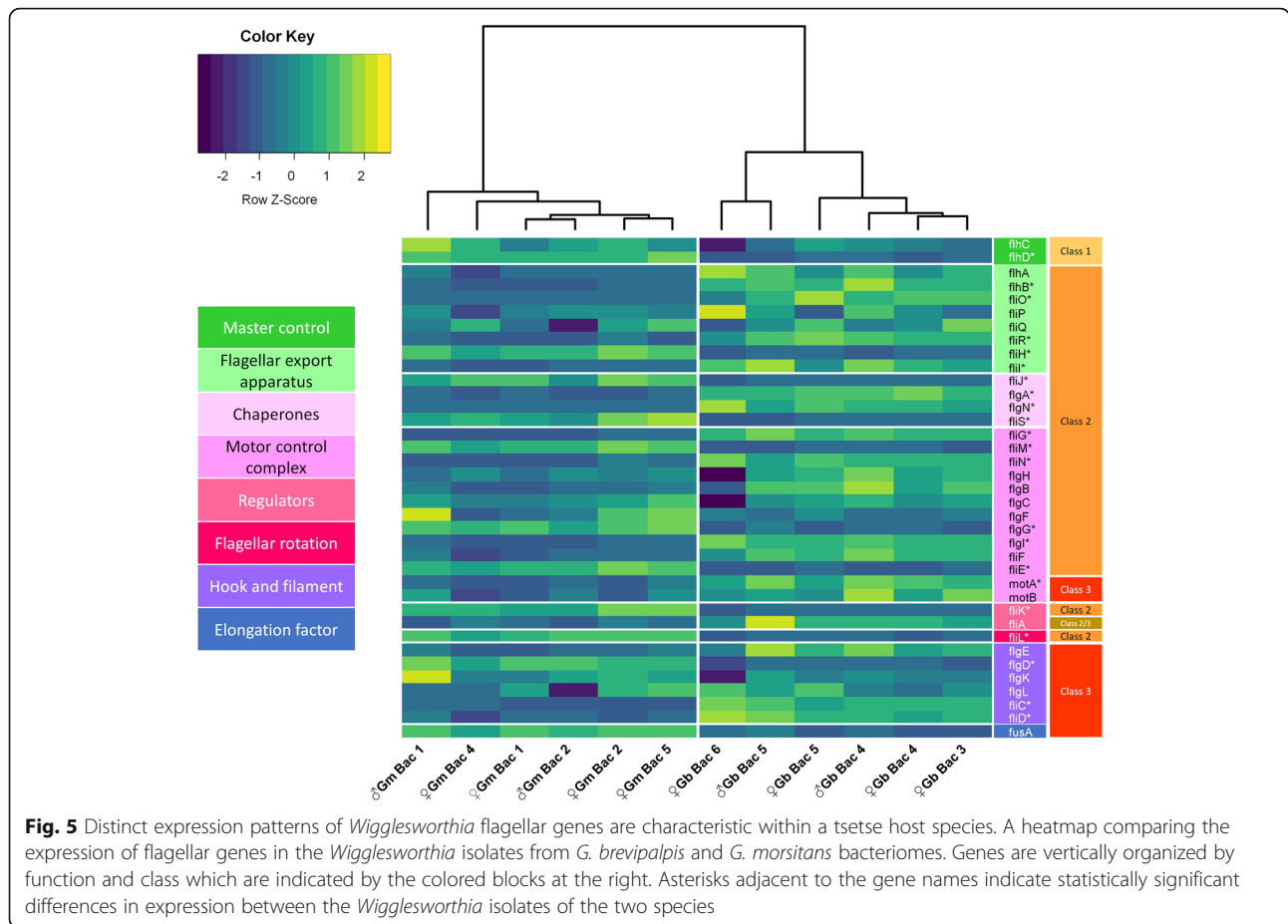
Fig. 5 Distinct expression patterns of *Wigglesworthia* flagellar genes are characteristic within a tsetse host species. A heatmap comparing the expression of flagellar genes in the *Wigglesworthia* isolates from *G. brevipalpis* and *G. morsitans* bacteriomes. Genes are vertically organized by function and class which are indicated by the colored blocks at the right. Asterisks adjacent to the gene names indicate statistically significant differences in expression between the *Wigglesworthia* isolates of the two species
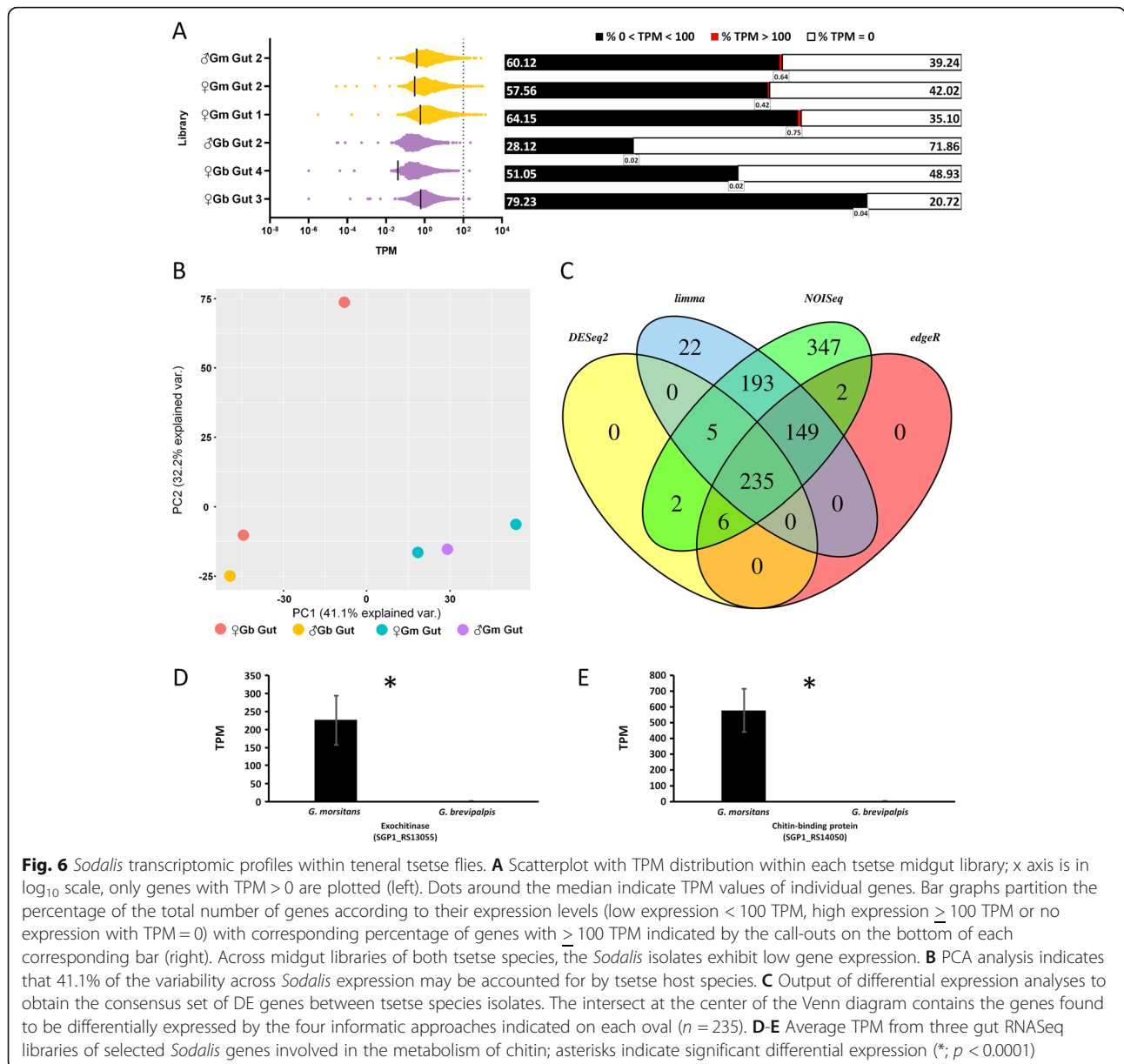
in *G. brevipalpis* isolates, nested *t*-test, $p = 0.6042$). Further, a PCA analysis shows that tsetse host species can account for 41% of gene expression variation (PC1) between the *Sodalis* isolates (Fig. 6B), with no clustering by host sex. The analysis of read counts via IDEAmex [52] identified 235 genes to be differentially expressed between *Sodalis* isolates of the two tsetse species (Fig. 6C, Additional file 5: Table S4). These genes represent a very small proportion of the total protein coding genes (5.2%), particularly when contrasted with the high fraction of orthologs differentially expressed between the two *Wigglesworthia* isolates (55%).

Sodalis exochitinase expression profile may contribute to tsetse vector competence

The higher susceptibility of teneral tsetse to trypanosome infections is thought to partly arise due to the immaturity of the peritrophic matrix (PM). The PM of adult tsetse is continuously produced by the cardia (characteristic of a Type II PM) and forms a protective semipermeable barrier within the intestinal tract by surrounding the blood bolus within an endoperitrophic space. Due to the PM being rich in chitin [53] coupled with the higher susceptibility of tsetse flies towards

trypanosome infection as tenerals [54], *Sodalis* genes encoding chitin-associated proteins may compromise PM integrity and thereby facilitate trypanosome infection [55–57]. Two genes involved in chitin-associated activities were among the most differentially expressed between the two *Sodalis* isolates (Fig. 6D and E). The genes encoding the predicted chitin-binding protein (NCBI Protein ID: BAE74790) and exochitinase (*chiA*; NCBI Protein ID: BAE74749) are significantly upregulated within the *Sodalis* isolate of *G. morsitans* relative to that of *G. brevipalpis* during the teneral host stage (Fig. 6D and E). The exochitinase gene is essential for *Sodalis* persistence within tsetse [58] as its chitinolytic activity produces *N*-acetyl-D-glucosamine (GlcNAc) which is the principal carbon source for this bacterium. Additionally, the GlcNAc monosaccharides inhibit antitrypanosomal lectins present in tsetse midgut [59] impeding their binding to trypanosome surface carbohydrates promoting parasite establishment [60]. The *Sodalis* exochitinase activity may also facilitate trypanosomes crossing into the ectoperitrophic space by disrupting the physical integrity of the PM.

Highly expressed Sodalis *genes contain a large proportion of DEGs.*

Medina Munoz *et al. BMC Genomics*   (2021) 22:400

Page 8 of 20

Fig. 6 *Sodalis* transcriptomic profiles within teneral tsetse flies. **A** Scatterplot with TPM distribution within each tsetse midgut library; x axis is in $\log_{10}$ scale, only genes with TPM > 0 are plotted (left). Dots around the median indicate TPM values of individual genes. Bar graphs partition the percentage of the total number of genes according to their expression levels (low expression < 100 TPM, high expression $\geq$ 100 TPM or no expression with TPM = 0) with corresponding percentage of genes with $\geq$ 100 TPM indicated by the call-outs on the bottom of each corresponding bar (right). Across midgut libraries of both tsetse species, the *Sodalis* isolates exhibit low gene expression. **B** PCA analysis indicates that 41.1% of the variability across *Sodalis* expression may be accounted for by tsetse host species. **C** Output of differential expression analyses to obtain the consensus set of DE genes between tsetse species isolates. The intersect at the center of the Venn diagram contains the genes found to be differentially expressed by the four informatic approaches indicated on each oval (n = 235). **D-E** Average TPM from three gut RNASeq libraries of selected *Sodalis* genes involved in the metabolism of chitin; asterisks indicate significant differential expression (*; $p < 0.0001$)

Interestingly, the highly expressed genes (TPM $\geq$ 100) in the *Sodalis* transcriptome contain also a disproportionately high number of DEGs (Fig. 7). As mentioned before, only about 5.2% of all *Sodalis* coding sequences are differentially expressed between the host species isolates; however, the subset of genes with TPM > 100 (Fig. 7) contains 55.3% of the total DEGs, which may indicate genes important for the *Sodalis*-tsetse symbiosis. For example, orthologs of type II toxin-antitoxin systems (SGP1_RS14115) along with a formation regulator BssS (SGP1_RS09085) facilitate biofilm formation [61, 62]; hypothetically this may support close proximity of *Sodalis* to the PM lining of the gut lumen,

where the bacterium may then deploy exochitinase (SGP1_RS13055), likely aided by lytic polysaccharide monooxygenase (SGP1_RS14050) and glycoside hydrolase family protein (SGP1_RS23580), in order to degrade chitin [63, 64] for access to GlcNAc as nutritional source. Although the main known role of chaperones is towards correcting misfolded proteins during stress response [65–68], they are also hypothesized to be important in mediating insect-bacteria interactions [69]. For example, *groES* constitutes one of the most abundant transcripts of the ant symbiont *Blochmannia* [70], *Wigglesworthia* in wild tsetse populations [71]) and *Buchnera* within aphids [72].
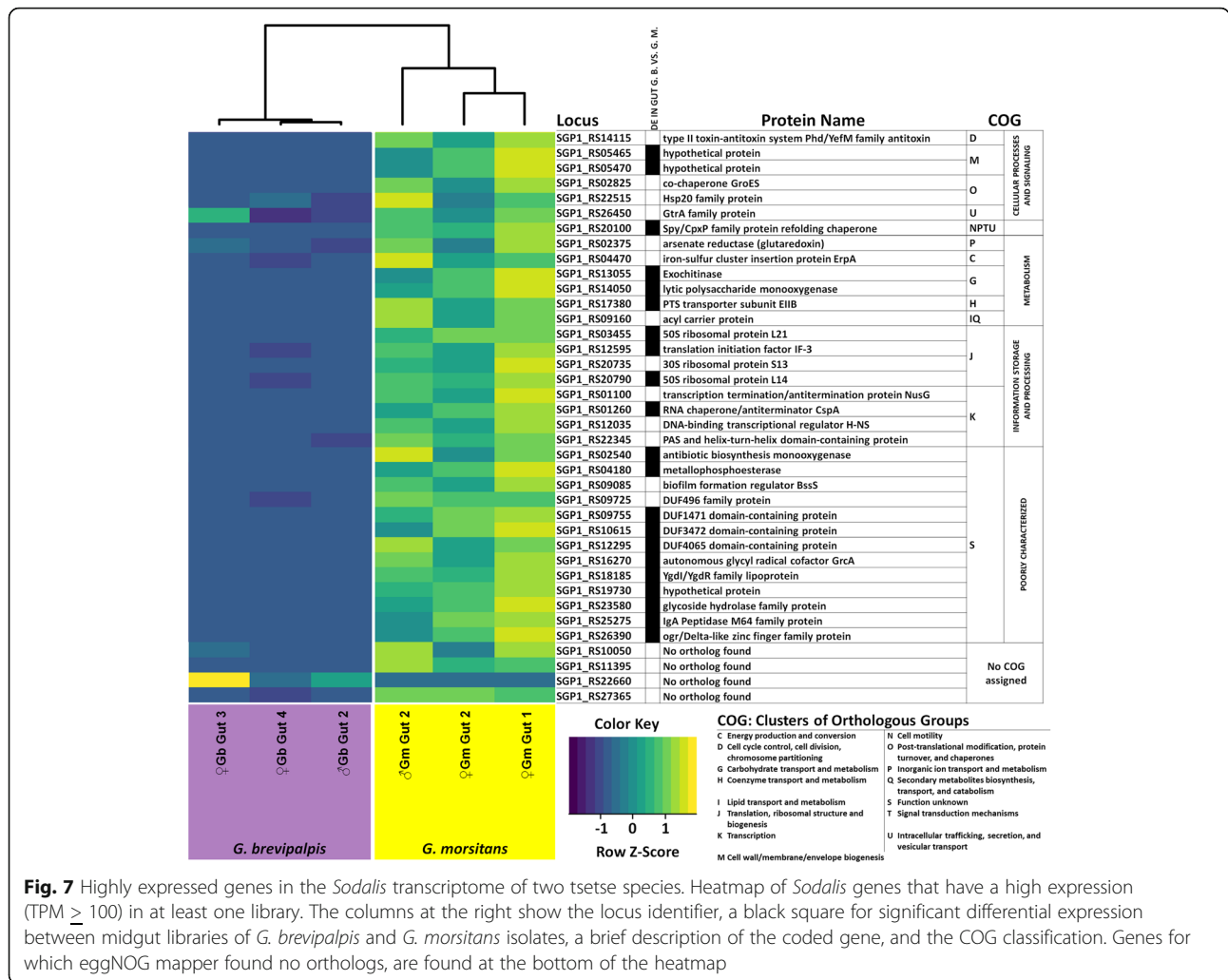
Fig. 7 Highly expressed genes in the *Sodalis* transcriptome of two tsetse species. Heatmap of *Sodalis* genes that have a high expression (TPM ≥ 100) in at least one library. The columns at the right show the locus identifier, a black square for significant differential expression between midgut libraries of *G. brevipalpis* and *G. morsitans* isolates, a brief description of the coded gene, and the COG classification. Genes for which eggNOG mapper found no orthologs, are found at the bottom of the heatmap

Tsetse fly based analyses

Multiple studies have demonstrated that trypanosome and bacterial infections in tsetse flies involve the expression of antimicrobial peptides (AMPs), reactive oxygen species (ROS) and other key anti-pathogen genes [73–77]. Yet, little has been done to contrast immunological profiles at specific developmental timepoints between tsetse species and integrating with the activity of the core microbiota. To complement the characterization of the symbiont transcriptomes, and to obtain a holistic picture of the genomic interplay within tsetse that may enhance our understanding of interspecific differences in host traits, we also examined the transcriptomes of *G. brevipalpis* and *G. morsitans*.

Tsetse transcriptomic profiles show a distinct clustering by species and tissues

If a tsetse gene was expressed, it likely had a transcription level of < 100 TPM (Fig. 8A). Only a relatively small proportion (~ 8.2%) of the total gene count is highly expressed (TPM ≥ 100); however, this number is significantly larger in the *G. brevipalpis* isolates [bacteriomes = 435.7 ± 24.65 (*n* = 6 libraries); midguts = 1197 ± 8.212 (*n* = 3 libraries)] when compared to *G. morsitans* [bacteriomes = 183.2 ± 10.45 (n = 6 libraries); midguts = 1007 ± 7.265 (n = 3 libraries)] in both bacteriomes and midgut libraries (*t*-test, *p* < 0.0001). When looking at specific groups of libraries, for genes with ≥ 100 TPM within midguts, the average TPM does not differ between the two tsetse host species [685.0 ± 44.69 (*n* = 3590 genes) in *G. brevipalpis* vs. 737.6 ± 70.59 (*n* = 3020 genes) in *G. morsitans* isolates, nested *t*-test, *p* = 0.5158]. However, when comparing genes with ≥ 100 TPM within bacteriome libraries, the average TPM is significantly higher for *G. brevipalpis* (overall average TPM ± SEM of 675.9 ± 52.98, *n* = 2614 genes) in comparison to *G. morsitans* (465.1 ± 25.45, *n* = 1099 genes, nested *t*-test, *p* = 0.0115). PCA analysis shows that 55.4% of the variability across tsetse gene expression is explained by tsetse tissue (PC1), while 23.8% of the variability may be accounted
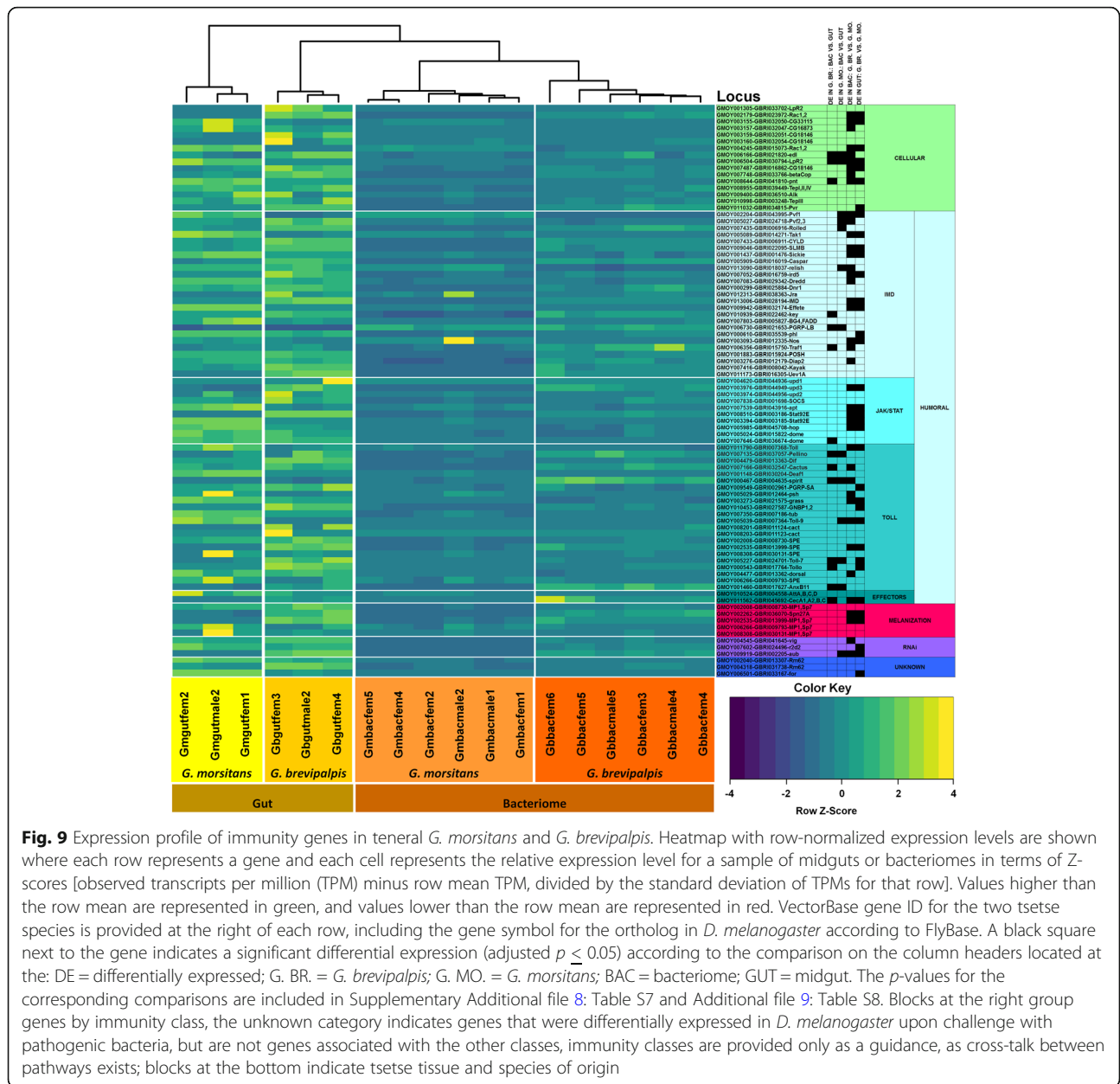
Fig. 8 Transcriptomic profiles of tsetse genes within teneral tsetse flies. **A** Scatterplot with TPM distribution within each tsetse library; x axis is in log$_{10}$ scale, only genes with TPM > 0 are plotted (left). The median is indicated with a vertical black line, TPM of 100 is indicated with dotted line; bar graphs partition the percentage of the total number of genes according to their expression levels (low expression < 100 TPM, high expression ≥ 100 TPM or no expression with TPM = 0). The call-outs on the side of each corresponding bar (right) indicate the percentage of genes with ≥ 100 TPM. Black dots at the left and right indicate the median and the cutoff of 100 TPM, respectively. **B** PCA analysis indicates that 55.4% of the variability across tsetse expression may be accounted for by tsetse tissue (PC1), while 23.8% of the variability may be accounted for by tsetse species (PC2). **C** Output of IDEAmex which depicts DE genes within the bacteriomes of tsetse species isolates; the intersect at the center of the Venn diagram contains the tsetse genes found to be differentially expressed by the four informatic approaches as identified adjacent to each oval; n = 3246 DE genes. **D** Output of IDEAmex which depicts DE genes within the midguts of tsetse species isolates; the intersect at the center of the Venn diagram contains the tsetse genes found to be differentially expressed by the four informatic approaches as identified adjacent to each oval; n = 3134 DE genes

for by tsetse species (PC2, Fig. 8B), with no clustering by fly sex. Differential expression analyses via IDEAmex shows 3246 DEG between *G. brevipalpis* and *G. morsitans* in bacteriome libraries (Fig. 8C, Additional file 6: Table S5) and 3134 DEG between *G. brevipalpis* and *G. morsitans* in midgut libraries (Fig. 8D, Additional file 7: Tables S6).

Tsetse species differ in the expression of immunity related pathways as tenerals

Given that immunity plays an essential role in mediating a spectrum of host-microbe interactions [78, 79], we compared the immunological transcriptomes of *G. brevipalpis* and *G. morsitans* bacteriomes and midguts. Immunity-related genes were identified via orthology

with *D. melanogaster* [80] and interspecies comparison performed [81]. These genes are critical components of various immunological responses including cellular, humoral, melanization and RNAi pathways (as identified in [80]). A heat map of immunity gene expression demonstrates clustering that distinguishes guts from bacteriomes and also by tsetse species (Fig. 9, Additional files 8 and 9: Tables S7 and S8 respectively). On average, genes involved in the various immunological mechanisms, are upregulated in midguts relative to the bacteriomes supporting the immunopermissive space of the bacteriome serving to protect essential *Wigglesworthia* symbionts (37.61 ± 3.442 TPM in midguts vs. 15.88 ± 1.154 TPM in bacteriomes; n = 86, nested t-test, p < 0.0001). This immune tolerance is further

Fig. 9 Expression profile of immunity genes in teneral *G. morsitans* and *G. brevipalpis*. Heatmap with row-normalized expression levels are shown where each row represents a gene and each cell represents the relative expression level for a sample of midguts or bacteriomes in terms of Z-scores [observed transcripts per million (TPM) minus row mean TPM, divided by the standard deviation of TPMs for that row]. Values higher than the row mean are represented in green, and values lower than the row mean are represented in red. VectorBase gene ID for the two tsetse species is provided at the right of each row, including the gene symbol for the ortholog in *D. melanogaster* according to FlyBase. A black square next to the gene indicates a significant differential expression (adjusted $p \leq 0.05$) according to the comparison on the column headers located at the: DE = differentially expressed; G. BR. = *G. brevipalpis*; G. MO. = *G. morsitans*; BAC = bacteriome; GUT = midgut. The *p*-values for the corresponding comparisons are included in Supplementary Additional file 8: Table S7 and Additional file 9: Table S8. Blocks at the right group genes by immunity class, the unknown category indicates genes that were differentially expressed in *D. melanogaster* upon challenge with pathogenic bacteria, but are not genes associated with the other classes, immunity classes are provided only as a guidance, as cross-talk between pathways exists; blocks at the bottom indicate tsetse tissue and species of origin

exemplified by the expression of *pgrp-lb* which is significantly higher within the bacteriomes of both tsetse species relative to midguts. Within the bacteriome the peptidoglycan recognition protein-LB (PGRP-LB) scavenges peptidoglycan released during *Wigglesworthia* cell division, thus preventing the activation of the hostile IMD pathway and offering symbiont protection [36, 82].

The cellular immunity category contains the only DEG identified between intraspecific bacteriomes and midguts and also between tsetse species, *LpR2* (GMOY006504, GBRI030794). *LpR2* has a significantly higher expression in the midgut libraries of both tsetse species when compared to the corresponding bacteriome expression levels.

Interspecies comparisons also indicate that *LpR2* is significantly upregulated within *G. morsitans* midguts and bacteriomes, when compared to the corresponding *G. brevipalpis* libraries. *LpR2* encodes a lipophorin receptor involved in the regulation of the Toll pathway [83]. An additional cellular immunity gene expressed significantly higher in *G. brevipalpis* midgut libraries was *pvr* which coordinates immunity responses through the inhibition of humoral immunity while stimulating hemocyte distribution, an early event in cellular immunity [84].

Humoral immunity pathways, such as Toll and Imd, are generally activated upon infection by Gram-positive and Gram-negative bacteria, respectively, and lead to the

production of distinct sets of AMPs, such as drosomycin, defensin and metchnikowin (Toll pathway) and attacin, cecropin and diptericin (Imd pathway) [85–89]. The *imd* gene is highly expressed both in *G. brevipalpis* midguts and bacteriomes. The gene *imd* is a strong regulator of antimicrobial responses against invading Gram-negative bacteria by inducing the expression of transcripts that encode antimicrobial effector proteins upon the recognition of specific microbial-associated molecular patterns [90], which is also consistent with the higher expression of *cecropin* within *G. brevipalpis* midguts [91]. A higher expression of *cecropin* may confer greater protection to *G. brevipalpis* against trypanosome infections, as products of orthologs in other insect species have killing activity against the related *Trypanosoma cruzi* [92, 93].

Interestingly, *effete* is highly expressed within midguts, but it is particularly higher in *G. morsitans*. The *effete* protein (*Ubc5*) mediates the polyubiquitination of IMD, leading to its degradation within the proteasome [94] which aligns with a lower *imd* expression. UevA joins *Ubc5* in the polyubiquitination of IMD [94]; *Uev1A* is highly expressed across libraries, however its expression is not significantly different between any of the library comparisons. These results may suggest a decreased potency of the IMD pathway in *G. morsitans* in the teneral state, which merits further investigation.

Within the Toll pathway, *G. brevipalpis* bacteriomes have a higher abundance of *spirit* transcripts in comparison to *G. morsitans*. The serine protease *Spirit* functions as a processing enzyme for the cytokine-like molecule Spätzle [95], a required initial step towards the activation of the Toll receptor for countering Gram-positive and fungal pathogens. The Toll pathway inhibitors *pellino* and *cactus* are upregulated in *G. brevipalpis* bacteriomes when compared to midguts. The proteins *Pellino* [96] and *Cactus* [97]) inhibit the Toll pathway by impeding signal transduction from the cell surface [96]) and by decreasing transcription of antimicrobial peptide coding genes [98], respectively. This may suggest that *G. brevipalpis* flies offer a more permissive microenvironment for bacterial growth in their bacteriomes. Additionally, *annexin B11* (*AnxB11*) was also upregulated in the bacteriomes of both species. Mammalian annexins are inhibitors of immune responses, where they suppress inflammatory responses during apoptosis [99, 100]. This observation further supports a more tolerant microenvironment in both tsetse species *G. brevipalpis* and *G. morsitans* likely aimed to sustain high bacterial densities within bacteriomes when compared to midguts.

Overall, genes involved in the JAK/STAT pathway show a low expression (TPM < 100) at the teneral stage [12.62 ± 1.401 TPM in JAK/STAT genes (n = 10) vs. 23.12 ± 1.405 TPM in all immune-related genes (n = 86); nested t-test, p = 0.0403], however, comparative

functional studies will be needed to validate differences in activation of the pathway between tsetse species. The cytokine *Upd3* mediates cellular immune response and is an activator of the JAK/STAT pathway [101]. Expression levels for *upd3* are consistently higher in *G. brevipalpis* bacteriome and midgut libraries relative to *G. morsitans*. Concurrently, the orthologs of the Signal-transducer and activator of transcription protein at 92E (*stat92E*) (GMOY008510 and GBRI003186) exhibit a higher expression in *G. brevipalpis* midguts. Upon pathogenic bacterial infection, Stat92E translocates to the nuclei of fat body cells where it drives antimicrobial peptide expression [102]. In contrast, a feedback inhibitor of the JAK/STAT pathway *Apontic* (*apt*) [103]) has relatively lower expression within *G. brevipalpis* libraries. It is puzzling that the ortholog set of the *stat92E* gene (GMOY003394 and GBRI003185) shows an opposite pattern of expression, where these have a higher level in *G. morsitans* bacteriomes and midguts when compared to the corresponding libraries within *G. brevipalpis*. This divergence in expression highlights the importance of functional characterization of genes across tsetse species.

In the melanization branch of immunity, both the Melanization Protease 1 [104] and Serine protease 7 (MP1-Sp7) may catalyze the initial activation of the pathway by cleaving the prophenoloxidase (PPO) zymogen to its active form phenoloxidase [105]. The serpin Spn27A is a serine protease inhibitor that negatively regulates melanization to limit the response to only the site of injury or infection preventing self-harm from excessive induction [106]. Interestingly, *G. brevipalpis* exhibits a simultaneously higher transcript abundance of both MP1-Sp7 (ortholog pair GMOY002535-GBRI013999) and Spn27A, which may allow this fly species to be poised to better respond should a pathogenic invasion occur and melanization required. Four tsetse ortholog pairs mapped to MP1-Sp7 in *D. melanogaster*, but the ortholog pairs GMOY002008-GBRI008730, GMOY006266-GBRI009793, and GMOY008308-GBRI030131 do not show a significant differential expression between tsetse species.

Discussion

Simultaneous host and symbiont examination enhances the understanding of their integrative biology

RNAseq enables the parallel assessment of transcriptomes from co-occurring species (many of which are unable to be separated without compromising the vitality of one or more partners) such as hosts and their microbiota. This approach is especially suitable for the discovery of novel points of interaction in host-microbe symbioses and the synthesis of robust hypotheses pertaining to how traits may be generated through host/microbiota activities. In this study, our main objective

was to identify differentially expressed host and symbiont bacteria genes, between two tsetse fly species at the teneral stage, and to characterize how these transcriptional profiles may impact host biology including vector competence. Additionally, we hypothesized that symbionts (i.e. *Wigglesworthia*) with a lengthier host co-evolution would show more distinctions in gene expression between tsetse species than recently acquired symbionts (i.e. *Sodalis*) which likely represent a greater extent of host adaptation.

The tsetse fly is recognized as a valuable animal model for enhancing our understanding of host-microbe symbiosis while also having high public health and agricultural significance as the obligate vector of African trypanosomes. As tenerals, tsetse have the highest susceptibility to trypanosome infections [25, 54, 107–110] likely due to a compilation of low levels of anti-trypanosomal binding midgut lectins, depleted fat reserves following metamorphosis, and the immature structural integrity of the PM crucial for both physical containment of trypanosomes within the endoperitrophic space and also towards midgut epithelial immune regulation [53, 60, 111]. Furthermore, vector competence varies between tsetse species with members of the *Morsitans* subgenus, including *G. morsitans,* exhibiting higher susceptibility to trypanosome infections, while those of the *Fusca* subgenus, such as *G. brevipalpis*, are comparatively poor vectors [112–114]. A deeper understanding of the molecular interactions between tsetse and its endogenous microbiota and how these may facilitate the establishment of trypanosomes during the teneral stage offer pillars for the development of novel and specific vector control strategies.

We expected to find that genes and pathways promoting trypanosome infection as tenerals would be enriched in expression in *G. morsitans* relative to *G. brevipalpis* which would associate a characteristic transcriptome profile of tsetse and its symbionts with vector competence. These interspecific distinctions may be further studied to understand functional diversification and relevancy towards tsetse biology and ecology. For example, differential transcriptomic profiles may help distinguish degrees of reliance among symbionts and tsetse species, likely influenced by differences in symbiont history with tsetse, and help to identify targets for disruption of fly-symbiont interactions in the context of novel vector control strategies targeting critical aspects of symbiosis.

The *Wigglesworthia* symbiont is not only essential for tsetse fly biology, but also its interaction with other members of the microbiota including *Sodalis* [115] and trypanosomes [27, 31]. Intriguingly, between 72 and 83% of *Wigglesworthia* genes are expressed within teneral tsetse flies indicating high activity following adult metamorphosis. The acquisition of nutrients by

trypanosomes must be strategically orchestrated with host metabolism requiring a fine balance between obtaining sufficient nutrients to complete its lifecycle but not sacrificing tsetse fitness to the extent that transmission to a naïve vertebrate host is compromised. Trypanosomes are auxotrophs for metabolites that are also essential for the fly and are available in very low amounts within blood, such as B vitamins [116, 117]. For example, *G. brevipalpis* harbors a *Wigglesworthia* isolate incapable of folate (B9) biosynthesis for which trypanosomes are also deficient. The exogenous supplementation of the *G. brevipalpis* diet with folate makes this tsetse species significantly more permissive to trypanosome establishment [31], highlighting trypanosome reliance on symbiont generated nutrients for successful vector infection. Here, metabolic properties of both tsetse and its bacterial symbionts may play key roles in trypanosomal nutrition and may signal parasite developmental cues that may ultimately impact the outcome of an infection. Genes in this category include those directly involved in vitamin and cofactor biosynthesis, such as those in the COG category of "Coenzyme transport and metabolism". Furthermore, a higher expression of *Sodalis* genes involved in chitin metabolism, such as the high expression of exochitinase and chitin-binding protein observed in the *G. morsitans* isolates, may have direct effects on trypanosome-tsetse interactions facilitating parasite establishment via competitive interference with lectins, while compromising the physical robustness of the PM which is primarily constituted of peritrophins and chitin. Lectin abundance within the midgut increases with tsetse age [110], therefore the breakdown of chitin by *Sodalis* would be less damaging towards anti-trypanosomal efforts as tsetse age [118]. Furthermore, chitin-binding proteins are thought to work in conjunction with exochitinases facilitating cell adhesion to cellular targets (reviewed in [119]), and orthologs have been implicated as virulence factors in various bacterial infections (reviewed in [120]).

The transcriptomic profile of *Wigglesworthia* flagellar genes represents two different stages of flagella synthesis between the species isolates. Flagella facilitate fine-tune responses to environmental stimuli by bacteria [121] and their regulation and synthesis is metabolically expensive, stressing the importance of global and master regulators genes to control their expression. The master regulator *flhDC* is transcribed into a single mRNA and translated into two distinct proteins FlhD and FlhC which assemble into the hexameric complex $FlhD_4C_2$ [122]. This complex attaches to promoter regions upstream of class II flagellar genes [51] facilitating ribosomal recruitment and transcription. Given the temporal regulation of this cascade (reviewed in [123, 124]), the stability of $FlhD_4C_2$ is under tight control [125] with even the degradation of

the *flhDC* mRNA controlled by global regulators such as CsrA [126, 127]. However, as *csrA* orthologs are absent in *Wigglesworthia* genomes, the regulation of master control genes such as *flhDC* and the potential role tsetse may play towards mediating flagellar control deserves further investigation. With this understanding, our results which show contrasting levels of *flhDC* transcription by *Wigglesworthia* may be due to measuring gene expression at a single time point. Despite the great advantage of looking at all the genes present in a pathway simultaneously, RNASeq lacks the temporal resolution necessary to reflect the dynamic nature of regulatory processes. This temporal regulation may explain our results, as for example, the *G. morsitans* isolates expression pattern may represent a stage where the *flhDC* mRNA is being highly transcribed, but it has yet to be translated to go on and exert its activation on class II gene transcription. This would account for a comparatively high expression of *flhDC* operon while the class II genes are still at low levels. Conversely, the pattern observed in *G. brevipalpis* may reflect a stage further in the flagellar synthesis cascade where the *flhDC* mRNA was already transcribed and mostly degraded, while the corresponding protein complex is activating (or has already) the transcription of downstream class II flagellar genes. Further studies following the mRNA stability and protein levels of *flhDC* across tsetse development, specifically encompassing the late pupae to early adult transition, may clarify the diverging regulation dynamics behind these expression patterns with likely implications towards acquisition of the obligate *Wigglesworthia* critical for tsetse developmental biology. Inhibiting the vertical transmission of *Wigglesworthia* renders female progeny sterile and may serve as a novel angle for next generation pesticides with sole specificity to the tsetse fly.

The low number of read counts arising from *Sodalis* is not surprising given that the abundance of 16S rRNA sequences belonging to *Sodalis* is relatively low in *G. morsitans* [12, 128]. The small proportion of differentially expressed genes in *Sodalis*, in comparison to *Wigglesworthia*, may reflect differences in symbiont acquisition times where the ancient *Wigglesworthia* is particularly finetuned to its host species due to their extensive coevolutionary history. In contrast, *Sodalis* is in an incipient stage of co-diversification [129, 130]. Two other notable features of the *Sodalis* transcriptome include the significantly greater proportion of highly expressed genes ($\geq$ 100 TPM) within the *G. morsitans* midgut isolates and the wide range of *Sodalis* genes expressed within *G. brevipalpis* individuals.

Tsetse fly fitness requires a balanced interaction with its microbiome and the environment

Invertebrates depend on innate immunity to mediate responses to microorganisms. These interactions may antagonize an infection, neutralize a pathogen, or even establish permissive (i.e. tolerance) conditions, thus promoting beneficial symbioses [78]. Beneficial symbionts have been found to orchestrate a coordinated development of the host immune system that simultaneously allows persistence while still enabling protection from pathogens. For example, *Burkholderia* symbionts differentially suppress host immunity to allow persistence [131] while playing a critical role in the proper functioning of the immune system in the bean bug *Riptortus pedestris* [132]. Similarly, *Wigglesworthia* is essential for the development of the immune system in tsetse [33, 34], given that larvae reared in the absence of this symbiont exhibit compromised humoral and cellular immune responses to pathogens as adults [34, 35].

The transcriptomic profiles of immunity related genes we observe in tsetse midguts and bacteriomes are consistent with an attenuated response to symbionts, rather than with "on" or "off" states, as for example regarding humoral immunity. The Toll pathway generally responds to challenge by Gram-positive bacteria and fungi, while the IMD pathway is dedicated towards the surveillance and protection against Gram-negative bacteria and fungi [89]. Our bacteriome libraries indicate that the Toll pathway is inactive in tenerals, not surprising, given that the sequencing libraries were generated from laboratory reared teneral flies lacking exposure to exogenous bacterial challenges. However, the Imd pathway is active within tenerals, as may be expected given that the tsetse microbiota is dominated by the Gram-negative bacteria, *Sodalis* and *Wigglesworthia*. However, these natural infections represent a physiological challenge for the tsetse, particularly due to the essential nature of its association with *Wigglesworthia*. Tsetse appear to have evolved mechanisms to circumvent this challenge by differential activation of the pathway within bacteriomes versus midguts and through the targeted expression of genes such as *pgrp-lb*, which when translated scavenges peptidoglycan released during *Wiggleworthia* cell division, thus preventing the activation of the hostile IMD pathway [36] which would prove detrimental towards *Wigglesworthia*.

Interestingly, several observations support an enhanced readiness in *G. brevipalpis* to deter trypanosome infections as tenerals. Previous transcriptome analyses in tsetse have found that activators of both Imd and Toll pathways are present in the fat body [133] and are necessary to counteract trypanosome infections and decrease their density [134]. The *imd* gene is highly expressed in *G. brevipalpis* midguts along with a concomitant high expression of the antimicrobial peptide cecropin. The cecropin peptide has potent killing activity against the American trypanosome *Trypanosoma cruzi*, [92, 93], so similar deleterious effects against African

trypanosomes may thus be hypothesized. Additionally, genes for enzymes that initiate the melanization cascade (i.e. MP1-Sp7) and their corresponding inhibitors (Serpin 27A) both have a higher expression in *G. brevipalpis*, which may enable a faster deployment should a trypanosome infection occur, as phenoloxidases have been implicated in response to trypanosome challenges in other insects such as triatomines [135] and bumble bees [136]). Consistent with this rationale, a pre-activation of immunity via artificial bacterial challenge, enables tsetse to respond more efficiently to a trypanosome challenge [74]. It is plausible that tsetse species provide microenvironments with different degrees of biological hostility toward trypanosomes, which would translate into distinctions in vector competence.

The differential expression observed in the set of genes arising from tsetse and its symbionts becomes even more relevant, as these occur at the teneral state. It seems possible that following adult metamorphosis, *G. brevipalpis* is comparatively better suited than *G. morsitans* to counter trypanosome infections. These gene expression patterns warrant a deeper investigation. For example, selective knockdown of genes or their activity, through RNA interference or chemically mediated inhibition, followed by trypanosome challenge would provide avenues targeted at assessing the contribution of identified genes towards vector competence.

Conclusions
The tsetse fly, similar to other animals, has to balance protection against pathogens with the biological integration of its essential microbiome. Our results indicate that this equilibrium may be, at least partially, achieved via a comparative downregulation of immunity in the compartments that harbor essential symbionts (i.e. bacteriomes in tsetse) relative to a direct route for pathogen entry, such as the digestive tract. Here, we show that bacterial symbionts exhibit transcriptomic profiles that reflect the duration of their respective host co-evolutionary histories, with a high percentage of DEGs in the ancient *Wigglesworthia* and a significantly lower proportion of DEGs in the more recent *Sodalis* when comparing tsetse species isolates. Furthermore, observed differences in the metatranscriptomes of the two tsetse species considered, such as the putatively higher deployment of antimicrobial peptide cecropin by *G. brevipalpis*, or higher transcription of enzymes with predicted chitinolytic activity in the *Sodalis* isolate from *G. morsitans*, offer insight into mechanisms that may predispose tsetse species to trypanosome establishment.

Methods
Insect rearing
Glossina morsitans morsitans pupae were provided by the Institute of Zoology, Slovak Academy of Sciences (Bratislava, Slovakia), and *Glossina brevipalpis* pupae were supplied by the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture (Vienna, Austria). Pupae were maintained in the Department of Biology insectary at West Virginia University at 24 ± 1 °C with 55% relative humidity on a 12-h light/12-h dark schedule until adult eclosion. Teneral flies (newly emerged adults prior to blood meal acquisition) were collected < 24 h post emergence from pupae and sorted by sex. Flies included in this study were all trypanosome free.

Dissections and RNA extraction
Bacteriomes and intestinal tracts (flanked by the bacteriome and the Malpighian tubules) were microscopically dissected and placed in RNA*later* (Invitrogen, Carlsbad, CA) at − 20 °C. The bacteriomes or intestinal tracts of 20 teneral tsetse of each sex and species were pooled for one biological sample, resulting in a total of 18 biological samples used in our analyses. Bacteriomes and guts were homogenized and total RNA was extracted using a MasterPure RNA purification kit (Epicentre, Madison, WI) according to the manufacturer's protocol for tissue samples. DNA was removed from the RNA samples using a Turbo DNA-free kit (Ambion, Austin, TX) following the rigorous DNase treatment option. The RNA concentration was measured using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) with an Agilent 2000 Bioanalyzer RNA Nano chip used to validate RNA sample quality and integrity.

mRNA library preparation, sequencing, and genome alignment
Library preparation was performed at the WVU Genomics Core Facility by using 1 μg of total RNA and Ribo-Zero Gold Epidemiology kit (Illumina, San Diego, CA) following the manufacturers recommended protocol. Following cDNA synthesis, libraries were quantified via Qubit fluorometer with high sensitivity DNA reagent and run on an Agilent high sensitivity DNA chip to determine average library size. The libraries were pooled in equimolar amounts and sequenced using the Illumina HiSeq 1500 platform (2 by 51 bp) at Marshall University. Following sequencing, raw reads were postprocessed to remove Illumina adapter and barcode sequences.

FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) analysis, as implemented in MultiQC [137], was performed on the RNA-Seq data sets to validate read quality. Reads were mapped to the *Glossina morsitans* (GCA_001077435.1), *Glossina brevipalpis*

(GCA_000671755.1), *Wigglesworthia* (GCA_000008885.1) and *Sodalis glossinidius* (GCA_000010085.1) genomes using Salmon [29]. Gene transcription level was converted to Transcripts per Million (TPM) fragments mapped and used for statistical comparison of expression levels between libraries [138]. Statistically significant differences were accepted at $p < 0.05$ with adjusted *p*-values for multiple testing.

Principal Component Analysis (PCA)
PCAs comparing gene TPM values were performed with the prcomp package (version 3.6.3) in the R software suite. PCA excluded genes that lacked expression across all libraries. PCA plots were visualized using the R package ggbiplot.

Differential expression analyses
IDEAmex [52], which implements the R packages DESeq2 [139], edgeR [140], limma [141] and NOISeq [142], with an adjusted *p*-value cut-off of 0.05 was used to compare gene expression profiles between *Wigglesworthia* isolates, between *Sodalis* isolates and between tsetse species. To compare *Wigglesworthia* gene expression between sexes within a species, the mapped read counts were initially used as input for DESeq. Additionally, EggNOG mapper [143] was used to assign Clusters of Orthologous Groups (COG, [49]) to categorically summarize annotation data into specific biological categories that were enriched within specific libraries.

Validation of differential expression through qRT-PCR
A subset of genes identified to be differentially expressed between libraries was verified via quantitative reverse transcription PCR (qRT-PCR). For the validation of the differential expression of *Wigglesworthia* genes, biological samples that consisted of six individual bacteriomes from either tsetse sex were collected in RNAlater (ThermoFisher Scientific, Waltham, MA) following manufacturer's protocol. Total RNA was isolated using MasterPure™ RNA (Epicentre, Madison, WI) and treated with TURBO™ DNase (ThermoFisher Scientific, Waltham, MA) following the rigorous treatment protocol to remove contaminant DNA. Linearized plasmid standards used for the quantification of gene copy number were made for respective genes using the pGEM®-T Vector Systems (Promega, Madison, WI) according to manufacturer instructions. Table S1 includes a list of primers used for cloning and qPCR amplification. First-strand cDNA synthesis was performed with SuperScript™ II Reverse Transcriptase (ThermoFisher Scientific, Waltham, MA). Second-strand cDNA synthesis was performed with a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA) using the SsoFast™ PCR cocktail (Bio-Rad) and following the conditions used in

Additional file 10: Table S9. Three technical replicates were performed for each biological sample and averages obtained. The relative gene expression was determined for selected genes using the $2^{-\Delta\Delta Ct}$ method [144].

Immunity
Putative orthologs of *D. melanogaster* immunity genes [80] in tsetse were identified by using FlyBase (https://flybase.org). Orthologs between *G. morsitans* and *G. brevipalpis* were identified through VectorBase (https://vectorbase.org). The list of immune-related orthologs was validated and placed into pathways according to [81].

Graphs and statistical analyses
Heatmaps were generated with the 'heatmap.2 function' in the gplots R package with clustering by species/tissue type based on the distinct expression patterns between isolates. GraphPad Prism was used for statistical analyses with *p*-values < 0.05 considered statistically significant.

Abbreviations
AMPs: Antimicrobial peptides; COGs: Clusters of orthologous groups of proteins; DEGs: Differentially expressed genes; PCA: Principal component analysis; PM: Peritrophic matrix; PPO: Prophenoloxidase; ROS: Reactive oxygen species; TPM: Transcripts per Million

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07729-5.

Additional file 1: Fig. S1. Mean quality scores by position of the reads. **Fig. S2.** Read count per library. **Fig. S3.** Comparison of total reads and mapped reads between tsetse species libraries. **Fig. S4.** Within species comparison of highly expressed *Wigglesworthia* genes among two tsetse species isolates.

Additional file 2: Table S1. Differentially expressed *Wigglesworthia* genes between female and male *G. morsitans* isolates.

Additional file 3: Table S2. Upregulation and downregulation of differentially expressed genes between *Wigglesworthia* isolates.

Additional file 4: Table S3. Highly expressed *Sodalis* genes.

Additional file 5: Table S4. *p*-values of differentially expressed genes between *Sodalis* isolates.

Additional file 6: Table S5. *p*-values of differentially expressed genes between *G. brevipalpis* and *G. morsitans* in bacteriome libraries.

Additional file 7: Table S6. *p*-values of differentially expressed genes between *G. brevipalpis* and *G. morsitans* in midgut libraries.

Additional file 8: Table S7. *p*-values of differentially expressed genes between *G. brevipalpis* bacteriomes and midgut libraries.

Additional file 9: Table S8. *p*-values of differentially expressed genes between *G. morsitans* bacteriomes and midgut libraries.

Additional file 10: Table S9. Primers used for the validation of *Wigglesworthia* differential expression though qRT-PCR.

Medina Munoz *et al. BMC Genomics*   (2021) 22:400

Page 17 of 20

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biology, Eberly College of Arts and Sciences, West Virginia University, Morgantown, WV 26505, USA. [2]Department of Biology, Washington and Jefferson College, Washington, PA 15301, USA.

References
1. Simarro PP, Cecchi G, Franco JR, Paone M, Diarra A, Ruiz-Postigo JA, et al. Estimating and mapping the population at risk of sleeping sickness. PLoS Negl Trop Dis. 2012;6(10):e1859. https://doi.org/10.1371/journal.pntd.0001859.
2. Simarro PP, Jannin J, Cattand P. Eliminating human African trypanosomiasis: where do we stand and what comes next? PLoS Med. 2008;5(2):e55. https://doi.org/10.1371/journal.pmed.0050055.
3. Jordan AM. Control of tsetse flies (Diptera: Glossinidae) with the aid of attractants. J Am Mosq Control Assoc. 1995;11(2 Pt 2):249–55.
4. Hargrove JW, Omolo S, Msalilwa JSI, Fox B. Insecticide-treated cattle for tsetse control: the power and the problems. Med Vet Entomol. 2000;14(2):123–30. https://doi.org/10.1046/j.1365-2915.2000.00226.x.
5. Hargrove JW, Torr SJ, Kindness HM. Insecticide-treated cattle against tsetse (Diptera: Glossinidae): what governs success? Bull Entomol Res. 2003;93(3):203–17. https://doi.org/10.1079/BER2003234.
6. Benoit JB, et al. Adenotrophic viviparity in tsetse flies: potential for population control and as an insect model for lactation. Annu Rev Entomol. 2015;60:351–71.
7. Ma WC, Denlinger DL. Secretory discharge and microflora of milk gland in tsetse flies. Nature. 1974;247(5439):301–3. https://doi.org/10.1038/247301a0.
8. Attardo GM, Lohs C, Heddi A, Alam UH, Yildirim S, Aksoy S. Analysis of milk gland structure and function in Glossina morsitans: milk protein production, symbiont populations and fecundity. J Insect Physiol. 2008;51:1236–442.
9. Balmand S, Lohs C, Aksoy S, Heddi A. Tissue distribution and transmission routes for the tsetse fly endosymbionts. J Invertebr Pathol. 2013;112(Suppl):S116–22. https://doi.org/10.1016/j.jip.2012.04.002.
10. Aksoy S. Wigglesworthia gen nov. and Wigglesworthia glossinidia sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbiont of tsetse flies. Int J Syst Bacteriol. 1995;45(4):848–51. https://doi.org/10.1099/00207713-45-4-848.
11. Dale C, Maudlin I. Sodalis gen. Nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. Int J Syst Bacteriol. 1999;49:267–75.
12. Aksoy E, Telleria EL, Echodu R, Wu Y, Okedi LM, Weiss BL, et al. Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. Appl Environ Microbiol. 2014;80(14):4301–12. https://doi.org/10.1128/AEM.00079-14.
13. Lindh JM, Lehane MJ. The tsetse fly Glossina fuscipes fuscipes (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie Van Leeuwenhoek. 2011;99(3):711–20. https://doi.org/10.1007/s10482-010-9546-x.
14. Geiger A, Fardeau ML, Grebaut P, Vatunga G, Josénando T, Herder S, et al. First isolation of Enterobacter, Enterococcus, and Acinetobacter spp. as inhabitants of the tsetse fly (Glossina palpalis palpalis) midgut. Infect Genet Evol. 2009;9(6):1364–70. https://doi.org/10.1016/j.meegid.2009.09.013.
15. Chen XA, Song L, Aksoy S. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-asssociated endosymbiont Wigglesworthia glossinidia. J Mol Evol. 1999;48(1):49–58. https://doi.org/10.1007/PL00006444.
16. Aksoy S, Pourhosseini AA, Chow A. Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to Enterobacteriaceae. Insect Mol Biol. 1995;4(1):15–22. https://doi.org/10.1111/j.1365-2583.1995.tb00003.x.
17. Wang JW, et al. Interactions between mutualist Wigglesworthia and tsetse peptidoglycan recognition protein (Pgrp-Lb) influence trypanosome transmission. Am J Trop Med Hyg. 2009;81(5):291.
18. Cheng QaSA. Tissue tropism, transmission, and expression of foreign genes in vivo in midgut symbionts of tsetse flies. Insect Mol Biol. 1999;8(1):125–32.
19. Dennis JW, Durkin SM, Horsley Downie JE, Hamill LC, Anderson NE, MacLeod ET. Sodalis glossinidius prevalence and trypanosome presence in tsetse from Luambe National Park, Zambia. Parasit Vectors. 2014;7(1):378. https://doi.org/10.1186/1756-3305-7-378.
20. Farikou O, Njiokou F, Mdiba Mdiba JA, Njitchouang GR, Djeunga HN, Asonganyi T, et al. Tripartite interactions between tsetse flies, Sodalis glossinidius and trypanosomes- an epidemiological approach in two historical human African trypanosomiasis foci in Cameroon. Infect Genet Evol. 2010;10:115–21.
21. Kame-Ngasse GI, Njiokou F, Melachio-Tanekou TT, Farikou O, Simo G, Geiger A. Prevalence of symbionts and trypanosome infections in tsetse flies of two villages of the "Faro and Déo" division of the Adamawa region of Cameroon. BMC Microbiol. 2018;18(Suppl 1):159. https://doi.org/10.1186/s12866-018-1286-5.
22. Simo G, Kanté ST, Madinga J, Kame G, Farikou O, Ilombe G, et al. Molecular identification of Wolbachia and Sodalis glossinidius in the midgut of Glossina fuscipes quanzensis from the Democratic Republic of Congo. Parasite. 2019;26:5. https://doi.org/10.1051/parasite/2019005.
23. Tsagmo Ngoune JM, Reveillaud J, Sempere G, Njiokou F, Melachio TT, Abate L, et al. The composition and abundance of bacterial communities residing in the gut of Glossina palpalis palpalis captured in two sites of southern Cameroon. Parasit Vectors. 2019;12(1):151. https://doi.org/10.1186/s13071-019-3402-2.
24. Trappeniers K, Matetovici I, van den Abbeele J, de Vooght L. The tsetse fly displays an attenuated immune response to its secondary symbiont. Front Microbiol. 2019;10:1650. https://doi.org/10.3389/fmicb.2019.01650.
25. Haines LR. Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection. Front Cell Infect Microbiol. 2013;3:84.
26. Snyder AK, McLain C, Rio RV. The tsetse fly obligate mutualist Wigglesworthia morsitans alters gene expression and population density via exogenous nutrient provisioning. Appl Environ Microbiol. 2012;78(21):7792–7. https://doi.org/10.1128/AEM.02052-12.
27. Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. Appl Environ Microbiol. 2014;80(18):5844–53. https://doi.org/10.1128/AEM.01150-14.
28. Snyder AK, Rio RV. "Wigglesworthia morsitans" Folate (vitamin B9) biosynthesis contributes to tsetse host fitness. Appl Environ Microbiol. 2015;81(16):5375–86. https://doi.org/10.1128/AEM.00553-15.

29. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017; 14(4):417–9. https://doi.org/10.1038/nmeth.4197.

30. Rio RV, et al. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. Proc Biol Sci. 2006;273(1588): 805–14.

31. Rio RVM, et al. Mutualist-provisioned resources impact vector competency. mBio. 2019;10(3):e00018.

32. Benoit JB, Vigneron A, Broderick NA, Wu Y, Sun JS, Carlson JR, et al. Symbiont-induced odorant binding proteins mediate insect host hematopoiesis. Elife. 2017;6. https://doi.org/10.7554/eLife.19535.

33. Weiss BL, Wang J, Aksoy S. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol. 2011;9(5):e1000619. https://doi.org/10.1371/journal.pbio.1000619.

34. Weiss BL, Maltz M, Aksoy S. Obligate symbionts activate immune system development in the tsetse fly. J Immunol. 2012;188(7):3395–403. https://doi. org/10.4049/jimmunol.1103691.

35. Weiss BL, Wang J, Maltz MA, Wu Y, Aksoy S. Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. PLoS Pathog. 2013;9(4):e1003318. https://doi.org/1 0.1371/journal.ppat.1003318.

36. Wang JW, Aksoy S. PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. Proc Natl Acad Sci U S A. 2012;109(26):10552–7. https://doi.org/10.1073/pnas.111 6431109.

37. Yang CP, Fu CC, Sugino K, Liu Z, Ren Q, Liu LY, et al. Transcriptomes of lineage-specific Drosophila neuroblasts profiled by genetic targeting and robotic sorting. Development. 2016;143(3):411–21. https://doi.org/10.1242/ dev.129163.

38. Yin W, Song Y, Chang X. Single-cell RNA-Seq analysis identifies a noncoding. J Biol Chem. 2019;294(1):290–8.

39. Macrander J, Panda J, Janies D, Daly M, Reitzel AM. Venomix: a simple bioinformatic pipeline for identifying and characterizing toxin gene candidates from transcriptomic data. PeerJ. 2018;6:e5361. https://doi.org/10. 7717/peerj.5361.

40. John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM. Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. Plant Physiol. 2014;165(1):62–75. https://doi.org/10.1104/pp.114.23 8667.

41. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. https://doi.org/10.1186/gb-2010-11-10-r106.

42. Rio RV, et al. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont Wigglesworthia. MBio. 2012;3(1):e00240.

43. Ibuki T, Imada K, Minamino T, Kato T, Miyata T, Namba K. Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. Nat Struct Mol Biol. 2011;18(3):277–82. https://doi.org/10.103 8/nsmb.1977.

44. Tso JY, Zalkin H, van Cleemput M, Yanofsky C, Smith JM. Nucleotide sequence of Escherichia coli purF and deduced amino acid sequence of glutamine phosphoribosylpyrophosphate amidotransferase. J Biol Chem. 1982;257(7):3525–31. https://doi.org/10.1016/S0021-9258(18)34810-5.

45. Sampei G, Mizobuchi K. Nucleotide sequence of the Escherichia coli purF gene encoding amidophosphoribosyltransferase for de novo purine nucleotide synthesis. Nucleic Acids Res. 1988;16(17):8717. https://doi.org/1 0.1093/nar/16.17.8717.

46. Missiakas D, Schwager F, Betton JM, Georgopoulos C, Raina S. Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in Escherichia coli. EMBO J. 1996;15(24): 6899–909. https://doi.org/10.1002/j.1460-2075.1996.tb01082.x.

47. Rohrwild M, Coux O, Huang HC, Moerschell RP, Yoo SJ, Seol JH, et al. HslV-HslU: a novel ATP-dependent protease complex in Escherichia coli related to the eukaryotic proteasome. Proc Natl Acad Sci U S A. 1996;93(12):5808–13. https://doi.org/10.1073/pnas.93.12.5808.

48. Seong IS, Oh JY, Lee JW, Tanaka K, Chung CH. The HslU ATPase acts as a molecular chaperone in prevention of aggregation of SulA, an inhibitor of cell division in Escherichia coli. FEBS Lett. 2000;477(3):224–9. https://doi. org/10.1016/S0014-5793(00)01808-1.

49. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997;278(5338):631–7. https://doi.org/10.1126/science.278. 5338.631.

50. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, et al. Genome sequence of the endocellular obligate symbiont of tsetse, Wigglesworthia glossinidia. Nat Genet. 2002;32(2):402–7.

51. Liu X, Matsumura P. The FlhD/FlhC complex, a transcriptional activator of the Escherichia coli flagellar class II operons. J Bacteriol. 1994;176(23):7345–51.

52. Jiménez-Jacinto V, Sanchez-Flores A, Vega-Alvarado L. Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): a web server tool for integrated RNA-Seq data analysis. Front Genet. 2019;10:279. https://doi. org/10.3389/fgene.2019.00279.

53. Weiss BL, Savage AF, Griffith BC, Wu Y, Aksoy S. The peritrophic matrix mediates differential infection outcomes in the tsetse fly gut following challenge with commensal, pathogenic, and parasitic microbes. J Immunol. 2014;193(2):773–82. https://doi.org/10.4049/jimmunol.1400163.

54. Van Hoof L, Henrard C, Peel E. Influences modificatrices de la transmissibilite cyclique du Trypanosoma gambiense par Glossina palpalis. Ann Soc Belg Med Trop. 1937;17:249–72.

55. Maudlin I, Ellis DS. Association between intracellular rickettsial-like infections of midgut cells and susceptibility to trypanosome infection in Glossina spp. Z Parasitenkd. 1985;71(5):683–7.

56. Welburn SC, Arnold K, Maudlin I, Gooday GW. Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. Parasitology. 1993;107(Pt 2):141–5. https://doi.org/10.1017/S0031182 00006724X.

57. Rose C, Belmonte R, Armstrong SD, Molyneux G, Haines LR, Lehane MJ, et al. An investigation into the protein composition of the teneral Glossina morsitans morsitans peritrophic matrix. PLoS Negl Trop Dis. 2014;8(4):e2691. https://doi.org/10.1371/journal.pntd.0002691.

58. Dale C, Welburn SC. The endosymbionts of tsetse flies: manipulating host-parasite interactions. Int J Parasitol. 2001;31(5–6):628–31. https://doi.org/10.1 016/S0020-7519(01)00151-5.

59. Ibrahim EA, Ingram GA, Molyneux DH. Haemagglutinins and parasite agglutinins in haemolymph and gut of Glossina. Tropenmed Parasitol. 1984; 35(3):151–6.

60. Maudlin I, Welburn SC. Lectin mediated establishment of midgut infections of Trypanosoma congolense and Trypanosoma brucei in Glossina morsitans. Trop Med Parasitol. 1987;38(3):167–70.

61. Xu J, et al. Identification of three type II toxin-antitoxin systems in. Toxins (Basel). 2018;10(11):467.

62. Domka J, Lee J, Wood TK. YliH (BssR) and YceP (BssS) regulate Escherichia coli K-12 biofilm formation by influencing cell signaling. Appl Environ Microbiol. 2006;72(4):2449–59. https://doi.org/10.1128/AEM.72.4.2449-2459.2 006.

63. Courtade G, Aachmann FL. Chitin-active lytic polysaccharide Monooxygenases. Adv Exp Med Biol. 2019;1142:115–29. https://doi.org/10.1 007/978-981-13-7318-3_6.

64. Eijsink V, Hoell I, Vaaje-Kolstada G. Structure and function of enzymes acting on chitin and chitosan. Biotechnol Genet Eng Rev. 2010;27:331–66. https:// doi.org/10.1080/02648725.2010.10648156.

65. Arsène F, Tomoyasu T, Bukau B. The heat shock response of Escherichia coli. Int J Food Microbiol. 2000;55(1–3):3–9. https://doi.org/10.1016/S0168-1 605(00)00206-3.

66. Guest RL, Raivio TL. Role of the gram-negative envelope stress response in the presence of antimicrobial agents. Trends Microbiol. 2016;24(5):377–90. https://doi.org/10.1016/j.tim.2016.03.001.

67. Liu J, Wang M, Yi H, Liu M, Zhu D, Wu Y, et al. ATPase activity of GroEL is dependent on GroES and it is response for environmental stress in Riemerella anatipestifer. Microb Pathog. 2018;121:51–8. https://doi.org/10.1 016/j.micpath.2018.04.029.

68. Susin MF, Baldini RL, Gueiros-Filho F, Gomes SL. GroES/GroEL and DnaK/ DnaJ have distinct roles in stress responses and during cell cycle progression in Caulobacter crescentus. J Bacteriol. 2006;188(23):8044–53. https://doi.org/10.1128/JB.00824-06.

69. Kupper M, Gupta SK, Feldhaar H, Gross R. Versatile roles of the chaperonin GroEL in microorganism-insect interactions. FEMS Microbiol Lett. 2014; 353(1):1–10. https://doi.org/10.1111/1574-6968.12390.

70. Stoll S, Feldhaar H, Gross R. Transcriptional profiling of the endosymbiont Blochmannia floridanus during different developmental stages of its holometabolous ant host. Environ Microbiol. 2009;11(4):877–88. https://doi. org/10.1111/j.1462-2920.2008.01808.x.

71. Medina Munoz M, Pollio AR, White HL, Rio RVM. Into the wild: parallel transcriptomics of the tsetse-Wigglesworthia mutualism within Kenyan

populations. Genome Biol Evol. 2017;9(9):2276–91. https://doi.org/10.1093/gbe/evx175.

72. Fares MA, Barrio E, Sabater-Muñoz B, Moya A. The evolution of the heat-shock protein GroEL from Buchnera, the primary endosymbiont of aphids, is governed by positive selection. Mol Biol Evol. 2002;19(7):1162–70. https://doi.org/10.1093/oxfordjournals.molbev.a004174.

73. Hu C, Rio RVM, Medlock J, Haines LR, Nayduch D, Savage AF, et al. Infections with immunogenic trypanosomes reduce tsetse reproductive fitness: potential impact of different parasite strains on vector population structure. PLoS Negl Trop Dis. 2008;2(3):e192. https://doi.org/10.1371/journal.pntd.0000192.

74. Hao Z, Kasumba I, Lehane MJ, Gibson WC, Kwon J, Aksoy S. Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc Natl Acad Sci U S A. 2001;98(22):12648–53. https://doi.org/10.1073/pnas.221363798.

75. Hao Z, Kasumba I, Aksoy S. Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidae). Insect Biochem Mol Biol. 2003; 33:1155–64.

76. MacLeod ET, et al. Antioxidants promote establishment of trypanosome infections in tsetse. Parasitology. 2007;134(6):827–31. https://doi.org/10.1017/S0031182007002247.

77. Vigneron A, Aksoy E, Weiss BL, Bing X, Zhao X, Awuoche EO, et al. A fine-tuned vector-parasite dialogue in tsetse's cardia determines peritrophic matrix integrity and trypanosome transmission success. PLoS Pathog. 2018; 14(4):e1006972. https://doi.org/10.1371/journal.ppat.1006972.

78. Eleftherianos I, et al. Endosymbiotic bacteria in insects: guardians of the immune system? Front Physiol. 2013;4:46.

79. Medzhitov R, Schneider DS, Soares MP. Disease tolerance as a defense strategy. Science. 2012;335(6071):936–41. https://doi.org/10.1126/science.1214935.

80. Obbard DJ, Welch JJ, Kim KW, Jiggins FM. Quantifying adaptive evolution in the Drosophila immune system. PLoS Genet. 2009;5(10):e1000698. https://doi.org/10.1371/journal.pgen.1000698.

81. Matetovici I, Caljon G, Van Den Abbeele J. Tsetse fly tolerance to T. brucei infection: transcriptome analysis of trypanosome-associated changes in the tsetse fly salivary gland. BMC Genomics. 2016;17(1):971.

82. Wang J, Wu Y, Yang G, Aksoy S. Interactions between mutualist Wigglesworthia and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypansome transmission. Proc Natl Acad Sci U S A. 2009;106: 12134–8.

83. Soukup SF, Culi J, Gubb D. Uptake of the necrotic serpin in Drosophila melanogaster via the lipophorin receptor-1. PLoS Genet. 2009;5(6):e1000532. https://doi.org/10.1371/journal.pgen.1000532.

84. Tsuzuki S, Matsumoto H, Furihata S, Ryuda M, Tanaka H, Jae Sung E, et al. Switching between humoral and cellular immune responses in Drosophila is guided by the cytokine GBP. Nat Commun. 2014;5(1):4628. https://doi.org/10.1038/ncomms5628.

85. Hedengren M, Borge K, Hultmark D. Expression and evolution of the Drosophila attacin/diptericin gene family. Biochem Biophys Res Commun. 2000;279(2):574–81. https://doi.org/10.1006/bbrc.2000.3988.

86. Hultmark D, Engström A, Andersson K, Steiner H, Bennich H, Boman HG. Insect immunity. Attacins, a family of antibacterial proteins from Hyalophora cecropia. EMBO J. 1983;2(4):571–6. https://doi.org/10.1002/j.1460-2075.1983.tb01465.x.

87. Imler JL, Bulet P. Antimicrobial peptides in Drosophila: structures, activities and gene regulation. Chem Immunol Allergy. 2005;86:1–21. https://doi.org/10.1159/000086648.

88. Lye SH, Chtarbanova S. Drosophila as a model to study brain innate immunity in health and disease. Int J Mol Sci. 2018;19(12):3922.

89. Meister S, Koutsos AC, Christophides GK. The Plasmodium parasite--a 'new' challenge for insect innate immunity. Int J Parasitol. 2004;34(13–14):1473–82. https://doi.org/10.1016/j.ijpara.2004.10.004.

90. Jo YH, Patnaik BB, Hwang J, Park KB, Ko HJ, Kim CE, et al. Regulation of the expression of nine antimicrobial peptide genes by TmIMD confers resistance against gram-negative bacteria. Sci Rep. 2019;9(1):10138. https://doi.org/10.1038/s41598-019-46222-8.

91. Onfelt Tingvall T, Roos E, Engström Y. The imd gene is required for local Cecropin expression in Drosophila barrier epithelia. EMBO Rep. 2001;2(3): 239–43. https://doi.org/10.1093/embo-reports/kve048.

92. Durvasula RV, et al. Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria. Proc Natl Acad Sci U S A. 1997;94(7):3274–8.

93. Fieck A, Hurwitz I, Kang AS, Durvasula R. Trypanosoma cruzi: synergistic cytotoxicity of multiple amphipathic anti-microbial peptides to T. cruzi and potential bacterial hosts. Exp Parasitol. 2010;125(4):342–7. https://doi.org/10.1016/j.exppara.2010.02.016.

94. Chen L, Paquette N, Mamoor S, Rus F, Nandy A, Leszyk J, et al. Innate immune signaling in. J Biol Chem. 2017;292(21):8738–49. https://doi.org/10.1074/jbc.M117.788158.

95. Kambris Z, Brun S, Jang IH, Nam HJ, Romeo Y, Takahashi K, et al. Drosophila immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for toll activation. Curr Biol. 2006;16(8):808–13. https://doi.org/10.1016/j.cub.2006.03.020.

96. Ji S, Sun M, Zheng X, Li L, Sun L, Chen D, et al. Cell-surface localization of Pellino antagonizes toll-mediated innate immune signalling by controlling MyD88 turnover in Drosophila. Nat Commun. 2014;5(1):3458. https://doi.org/10.1038/ncomms4458.

97. Germani F, Hain D, Sternlicht D, Moreno E, Basler K. The toll pathway inhibits tissue growth and regulates cell fitness in an infection-dependent manner. Elife. 2018;7. https://doi.org/10.7554/eLife.39939.

98. Liu B, Zheng Y, Yin F, Yu J, Silverman N, Pan D. Toll receptor-mediated hippo signaling controls innate immunity in Drosophila. Cell. 2016;164(3): 406–19. https://doi.org/10.1016/j.cell.2015.12.029.

99. Weyd H, Abeler-Dörner L, Linke B, Mahr A, Jahndel V, Pfrang S, et al. Annexin A1 on the surface of early apoptotic cells suppresses CD8+ T cell immunity. PLoS One. 2013;8(4):e62449. https://doi.org/10.1371/journal.pone.0062449.

100. Bollinger AL, Bollinger T, Rupp J, Shima K, Gross N, Padayachy L, et al. Annexin V expression on CD4. Immunology. 2020;159(2):205–20. https://doi.org/10.1111/imm.13140.

101. Yang H, Kronhamn J, Ekström JO, Korkut GG, Hultmark D. JAK/STAT signaling in Drosophila muscles controls the cellular immune response against parasitoid infection. EMBO Rep. 2015;16(12):1664–72. https://doi.org/10.15252/embr.201540277.

102. Agaisse H, Petersen UM, Boutros M, Mathey-Prevot B, Perrimon N. Signaling role of hemocytes in Drosophila JAK/STAT-dependent response to septic injury. Dev Cell. 2003;5(3):441–50. https://doi.org/10.1016/S1534-5807(03)00244-2.

103. Starz-Gaiano M, Melani M, Wang X, Meinhardt H, Montell DJ. Feedback inhibition of Jak/STAT signaling by apontic is required to limit an invasive cell population. Dev Cell. 2008;14(5):726–38. https://doi.org/10.1016/j.devcel.2008.03.005.

104. Tang H, Kambris Z, Lemaitre B, Hashimoto C. Two proteases defining a melanization cascade in the immune system of Drosophila. J Biol Chem. 2006;281(38):28097–104. https://doi.org/10.1074/jbc.M601642200.

105. Castillejo-López C, Häcker U. The serine protease Sp7 is expressed in blood cells and regulates the melanization reaction in Drosophila. Biochem Biophys Res Commun. 2005;338(2):1075–82. https://doi.org/10.1016/j.bbrc.2005.10.042.

106. De Gregorio E, et al. An immune-responsive Serpin regulates the melanization cascade in Drosophila. Dev Cell. 2002;3(4):581–92. https://doi.org/10.1016/S1534-5807(02)00267-8.

107. Distelmans W, D'Haeseleer F, Kaufman L, Rousseeuw P. The susceptibility of Glossina palpalis palpalis at different ages to infection with Trypanosoma congolense. Ann Soc Belg Med Trop. 1982;62(1):41–7.

108. Otieno LH, Darji N, Onyango P, Mpanga E. Some observations on factors associated with the development of Trypanosoma brucei brucei infections in Glossina morsitans morsitans. Acta Trop. 1983;40(2):113–20.

109. Welburn SC, Maudlin I. The nature of the teneral state in Glossina and its role in the acquisition of trypanosome infection in tsetse. Ann Trop Med Parasitol. 1992;86(5):529–36. https://doi.org/10.1080/00034983.1992.11812703.

110. Lehane MJ, Msangi AR. Lectin and peritrophic membrane development in the gut of Glossina m.morsitans and a discussion of their role in protecting the fly against trypanosome infection. Med Vet Entomol. 1991;5(4):495–501. https://doi.org/10.1111/j.1365-2915.1991.tb00578.x.

111. Maudlin I, Welburn SC. The role of lectins and trypanosome genotype in the maturation of midgut infections in Glossina morsitans. Trop Med Parasitol. 1988;39(1):56–8.

112. Moloo SK, Sabwa CL, Kabata JM. Vector competence of Glossina pallidipes and G. morsitans centralis for Trypanosoma vivax, T. congolense and T. b. brucei. Acta Trop. 1992;51(3–4):271–80. https://doi.org/10.1016/0001-706X(92)90045-Y.

113. Motloang M, Masumu J, Mans B, van den Bossche P, Latif A. Vector competence of Glossina austeni and Glossina brevipalpis for Trypanosoma congolense in KwaZulu-Natal, South Africa. Onderstepoort J Vet Res. 2012; 79(1):E1–6. https://doi.org/10.4102/ojvr.v79i1.353.

114. Leak SGA. Tsetse biology and ecology, their role in the epidemiology and control of trypanosomes. New York: CABI publishing; 1999.

115. Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc Biol Sci. 2010;277(1692):2389–97.

116. Klein CC, Alves JMP, Serrano MG, Buck GA, Vasconcelos ATR, Sagot MF, et al. Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. PLoS One. 2013;8(11):e79786. https://doi.org/10.1371/journal.pone.0079786.

117. Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J. Comparative metabolism of free-living Bodo saltans and parasitic Trypanosomatids. J Eukaryot Microbiol. 2016;63(5):657–78. https://doi.org/10.1111/jeu.12315.

118. Rose C, et al. Trypanosoma brucei colonizes the tsetse gut via an immature peritrophic matrix in the proventriculus. Nat Microbiol. 2020;5:909.

119. Guillén D, Sánchez S, Rodríguez-Sanoja R. Carbohydrate-binding domains: multiplicity of biological roles. Appl Microbiol Biotechnol. 2010;85(5):1241–9. https://doi.org/10.1007/s00253-009-2331-y.

120. Frederiksen RF, Paspaliari DK, Larsen T, Storgaard BG, Larsen MH, Ingmer H, et al. Bacterial chitinases and chitin-binding proteins as virulence factors. Microbiology (Reading). 2013;159(Pt 5):833–47. https://doi.org/10.1099/mic.0.051839-0.

121. Rossi E, Paroni M, Landini P. Biofilm and motility in response to environmental and host-related signals in gram negative opportunistic pathogens. J Appl Microbiol. 2018;125(6):1587–602. https://doi.org/10.1111/jam.14089.

122. Wang S, Fleming RT, Westbrook EM, Matsumura P, McKay DB. Structure of the Escherichia coli FlhDC complex, a prokaryotic heteromeric regulator of transcription. J Mol Biol. 2006;355(4):798–808. https://doi.org/10.1016/j.jmb.2005.11.020.

123. Macnab RM. Genetics and biogenesis of bacterial flagella. Annu Rev Genet. 1992;26(1):131–58. https://doi.org/10.1146/annurev.ge.26.120192.001023.

124. Chevance FF, Hughes KT. Coordinating assembly of a bacterial macromolecular machine. Nat Rev Microbiol. 2008;6(6):455–65. https://doi.org/10.1038/nrmicro1887.

125. Claret L, Hughes C. Rapid turnover of FlhD and FlhC, the flagellar regulon transcriptional activator proteins, during Proteus swarming. J Bacteriol. 2000; 182(3):833–6. https://doi.org/10.1128/JB.182.3.833-836.2000.

126. Wei BL, et al. Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli. Mol Microbiol. 2001;40(1):245–56.

127. Yakhnin AV, Baker CS, Vakulskas CA, Yakhnin H, Berezin I, Romeo T, et al. CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. Mol Microbiol. 2013;87(4):851–66. https://doi.org/10.1111/mmi.12136.

128. Doudoumis V, Blow F, Saridaki A, Augustinos A, Dyer NA, Goodhead I, et al. Challenging the Wigglesworthia, Sodalis, Wolbachia symbiosis dogma in tsetse flies: Spiroplasma is present in both laboratory and natural populations. Sci Rep. 2017;7(1):4699. https://doi.org/10.1038/s41598-017-04740-3.

129. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, et al. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. Genome Res. 2006;16(2):149–56. https://doi.org/10.1101/gr.4106106.

130. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, et al. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. PLoS Genet. 2012;8(11):e1002990. https://doi.org/10.1371/journal.pgen.1002990.

131. Kim JK, Lee JB, Jang HA, Han YS, Fukatsu T, Lee BL. Understanding regulation of the host-mediated gut symbiont population and the symbiont-mediated host immunity in the Riptortus-Burkholderia symbiosis system. Dev Comp Immunol. 2016;64:75–81. https://doi.org/10.1016/j.dci.2016.01.005.

132. Kim JK, Lee JB, Huh YR, Jang HA, Kim CH, Yoo JW, et al. Burkholderia gut symbionts enhance the innate immunity of host Riptortus pedestris. Dev Comp Immunol. 2015;53(1):265–9. https://doi.org/10.1016/j.dci.2015.07.006.

133. Attardo GM, Strickler-Dinglasan P, Perkin SAH, Caler E, Bonaldo MF, Soares MB, et al. Analysis of fat body transcriptome from the adult tsetse fly, Glossina morsitans morsitans. Insect Mol Biol. 2006;15(4):411–24. https://doi.org/10.1111/j.1365-2583.2006.00649.x.

134. Hu C, Aksoy S. Innate immune responses regulate trypanosome parasite infection of the tsetse fly Glossina morsitans morsitans. Mol Microbiol. 2006; 60(5):1194–204. https://doi.org/10.1111/j.1365-2958.2006.05180.x.

135. González-Rete B, Salazar-Schettino PM, Bucio-Torres MI, Córdoba-Aguilar A, Cabrera-Bravo M. Activity of the prophenoloxidase system and survival of triatomines infected with different Trypanosoma cruzi strains under different temperatures: understanding Chagas disease in the face of climate change. Parasit Vectors. 2019;12(1):219. https://doi.org/10.1186/s13071-019-3477-9.

136. Brown MJ, Moret Y, Schmid-Hempel P. Activation of host constitutive immune defence by an intestinal trypanosome parasite of bumble bees. Parasitology. 2003;126(Pt 3):253–60. https://doi.org/10.1017/S0031182002002755.

137. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–8. https://doi.org/10.1093/bioinformatics/btw354.

138. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 2010; 26(4):493–500. https://doi.org/10.1093/bioinformatics/btp692.

139. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8.

140. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616.

141. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

142. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011;21(12):2213–23. https://doi.org/10.1101/gr.124321.111.

143. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14. https://doi.org/10.1093/nar/gky1085.

144. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. Methods. 2001;25(4):402–8. https://doi.org/10.1006/meth.2001.1262.

Publisher's Note