2021

Initial and Advanced Stages of Microbiota Establishment within the Tsetse Fly

Miguel Eduardo Medina Munoz
West Virginia University, mem0007@mix.wvu.edu

Initial and Advanced Stages of Microbiota Establishment within the Tsetse Fly

Miguel E. Medina Munoz

Dissertation submitted to
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Philosophiae Doctor (PhD)
in
Biology

Rita V.M. Rio, Ph.D., *Chair*
Geoffrey Attardo, Ph.D.
Timothy Driscoll, Ph.D.
Jennifer Gallagher, Ph.D.
Teiya Kijimoto, Ph.D.

Department of Biology

Morgantown, West Virginia, USA
2021

Keywords: symbiosis, tsetse, microbiota, *Sodalis*, *Wigglesworthia*, transcriptome, epigenetics, DNA methylation, quorum-sensing

ABSTRACT

Initial and Advanced Stages of Microbiota Establishment within the Tsetse Fly

Miguel E. Medina Munoz

Symbiosis is a long-term physical association between two or more different species and range in association from facultative parasitism to obligate mutualism, although little is known regarding the early evolutionary steps that lead to its establishment, particularly the integration that happens at the genetic level Tsetse flies are the vector of African trypanosomes, causative agents of Human and Animal African Trypanosomiases and they provide an ideal model for the study of both initial and advanced stages of symbiosis. Tsetse have a simple digestive tract microbiota primarily consisting of two bacteria; the ancient mutualist *Wigglesworthia glossinidia* and the recently acquired *Sodalis glossinidius*. This work presents a chronological study in evolutionary terms of the history of a microbial-insect association, using the tsetse and its symbionts as a model. First, I present concepts on symbiosis, general trends observed in bacterial symbionts, and include comments on tsetse biology, ecology and their native microbiota. Second, we focus on early evolutionary events that mediate the transition of an environmental bacterium to a symbiotic lifestyle. I show that quorum sensing virulence suppression plays an integral role in facilitating the establishment of *Sodalis*-allied symbionts in diverse insect hosts. This knowledge contributes to the understanding of the early evolutionary steps involved in the formation of insect-bacterial symbiosis. Further, despite having no established history of interaction with tsetse, the bacterium *Sodalis praecaptivus* can infect reproductive tissues, enabling vertical transmission through adenotrophic viviparity within a single host generation. This creates an option for the use a genetically modified strain of *S. praecaptivus* in the biocontrol of insect disease vectors via paratransgenesis. Third, we take a closer look at well-established symbionts: the tsetse flies and their endogenous microbiota. I characterize and compare the metatranscriptome of teneral *Glossina morsitans* (high vector competence) to that of *Glossina brevipalpis* (low vector competence). I show that the transcriptome of *Wigglesworthia* and *Sodalis* reflect differences in the extent of co-evolution with tsetse and identify molecular components and pathways that may contribute towards distinctions in vector competence between tsetse host species. Fourth, I present a theoretical mechanism, based on current knowledge, about a novel way in which microbial symbionts may be exerting control on insect host via epigenetics. It is known that some bacteria in insect symbioses have a folate provisioning role; also, that folate is essential for methylation reactions. However, little is known about how DNA methylation operates across insect taxa or how symbionts influence this process. I propose a link between the folate provisioning roles of the microbiota and the health of blood feeding insects, via the epigenetic mechanism of DNA methylation. We highlight key concepts to epigenetic in symbiosis and gaps in knowledge relevant to this interaction. Towards the end, we summarize common biological challenges relevant to the hypothesis and corresponding ways to address them. Lastly, we summarize the advantages and the potential of the combined approach reflected in this work to gain a comprehensive understanding of symbiosis evolution, its mechanisms at organismal level and the potential to impact epidemiology, providing helpful tools for novel vector control strategies.

This work is dedicated in part to two keystone figures in the field of symbiosis, the late Dr. Lynn Margulis, and Dr. Nancy Moran

Primum non nocere

"But to highlight the need for monitoring and forecasting is also to highlight the urgency of the problem and the discomfiting reality of how much remains unknown."
-David Quammen

PREFACE

Tsetse flies are the vector of African trypanosomes. Tsetse also have a simple digestive tract microbiota primarily consisting of two bacteria; the ancient mutualist *Wigglesworthia glossinidia* and the recently acquired *Sodalis glossinidius*. These symbionts have distinct association times and impacts towards tsetse biology. Despite the widely recognized importance of the microbiota towards host physiology, there remains the question of how the integration that coordinates tsetse and microbiota functions, enabling first establishment, and subsequently, persistence of an association. This knowledge is epidemiologically relevant because impeding critical tsetse-microbiota processes offers new targets for vector control.

The scientific premise for this research is that tsetse provides an ideal model for the study of both initial and advanced stages of symbiosis. The former exemplified by the newly acquired *Sodalis* and the latter by the ancient endosymbiont *Wigglesworthia*. Our long-term goal is to provide a mechanistic framework for how microbiota integrate into vector biology to coordinate activities that promote evolutionary fitness. Our central hypothesis is that mutualism matures from an environmentally acquired infection towards a co-evolved host-microbe interdependency. Attenuated virulence facilitates a stable relationship between the host and infectious agent, particularly prior to the association, yielding novel and beneficial collaborations. This hypothesis will be examined with the following aims:

Specific Aim 1: To characterize the significance of quorum sensing (QS) towards initiating tsetse symbiosis. A diverse range of insects harbor *Sodalis*-allied symbionts, yet little is known about the molecular mechanisms that initiate these interactions. *Sodalis praecaptivus* is an environmental bacterium amenable to genetic manipulation. Our working hypothesis is that *S. praecaptivus* is a free-living generalist that represents a progenitor-like *Sodalis* that uses QS to establish host infection enabling integration into tsetse life history. Wild type *S. praecaptivus* and QS mutants will be introduced into tsetse flies and survival, fecundity and vertical transmission assessed. These results will further our understanding of the molecular mechanisms involved in the initiation of symbioses.

Specific Aim 2: To characterize the *Wigglesworthia* transcriptome within teneral flies of varying vector competence. The teneral (i.e., newly eclosed adult) life stage has the highest susceptibility to trypanosome infections. Further, tsetse species vary in vector competence (i.e., the ability to acquire, maintain and transmit trypanosomes). The objective of this aim is to compare the metabolic contributions of *Wigglesworthia* within teneral hosts of different vector competence. RNASeq will be used to compare the *Wigglesworthia* transcriptome between tsetse species of high (*G. morsitans*) and low (*G. brevipalpis*) vector competence. Our working hypothesis is that *Wigglesworthia* isolates will demonstrate significant differences in expression profiles that translate into functional distinctions in host-symbiont interdependency, representing diversification of the association between tsetse species and likely modulating vector competence.

Specific Aim 3: To establish a link between the provisioning role of microbiota and the regulation of host genetic potential via epigenetics. We expect to draw support from literature for the impact of microbiota folate provisioning role on insect host DNA methylation. Folate is necessary for the generation of the donor of methyl groups, *S*-adenosylmethionine. We hypothesize that DNA methylation may be a mechanism for microbiota to control host gene expression, particularly in blood-feeding insects, with a diet restricted in B vitamins. We expect to provide the basic conceptual framework for the study of DNA methylation as a mediator between bacterial symbionts and blood-feeding insect hosts.

Host-microbiota interactions are essential for animal biology with the disruption of these associations offering novel targets for vector control. Importantly, this requires a fundamental understanding of symbiont establishment and ensuing coordination of host-microbiota activities.

Acknowledgements

I wish to thank my advisor, Dr. Rita V.M. Rio, for immense, invaluable and constant support during the years invested in this project. A giant source of knowledge, guidance and experience. Thanks to the members of my Committee for directions during the design and development of experiments, and for always being available and supportive. Thanks to M.S. Adam Pollio for his friendship and support during the several stages of my training, experimentation and writing. Also, for introducing me to the North American culture in all its aspects beyond Science. Thanks to Ph.D. candidate Mason Lee for thoughtful considerations and discussions on scientific matters in Biology and much more.

Thanks to Dr. Colin Dale and Dr. Shinichiro Enomoto, from The University of Utah, for providing bacterial strains and for the close collaboration during all the stages of the trials involving *S. praecaptivus*. Thanks to the WVU Genomics Core, specially to Niel Infante, for help and insights during initial processing of transcriptomic data. Thanks to Dr. Jim McGraw and Dr. Kenneth Ryan for priceless directions on statistical analysis.

I am immensely grateful to the students that collaborated in different projects as undergrads: Noah Spencer, whose help proved invaluable toward the advancement of the *S. praecaptivus* trials, and whose work greatly expanded the line of research on the *Wigglesworthia* genome in the lab. Afsoon Sabet, for providing initial insights into DNA methylation in the tsetse system. Caitlyn Brenner and Dylan Richmond for their dedicated help during the analysis and verification involved in the transcriptomic project. I thank our Lab technician, Ying Zhang, for her constant support during the long hours in the lab.

Thanks to Dr. Katrina Stewart, Dr. Cheryl Walton and Dr. Dana Hubert-Lima for training and encouraging me to develop teaching skills during the several semesters I had the privilege to serve as Teaching Assistant for the Department. I am infinitely thankful to all the Professors of my grad level courses at WVU whose lectures were a constant source of new ideas, motivation and wonder.

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

Symbiosis

Symbiosis is a fluid long-term physical association between two or more different species [1] and may range in association from facultative parasitism to obligate mutualism. Symbionts are host-associated microbiota (i.e., microbes living within or on a host constituted by different microbial taxa. The metagenome is the collective genomes of these microbial communities [2–6]. Members of the microbiota play significant roles in insect physiology (including metabolism and immunology), behavior, ecology, and evolution [7–10]. In host-microbe relations, one of the partners benefits from the association while the other receives no benefit, referred to as commensalism; while if both partners benefit, the symbiosis is called mutualism. The association is parasitic or pathogenic when one of the member's benefits while causing a detrimental effect to the other, for instance deteriorating its health, altering the resource allocation patterns, or decreasing its fitness.

Symbioses are established by different routes. For example, members of the microbiota may be acquired horizontally (yet selectively) from the assemblage of microorganisms present in the environment (like within water or soil) or they may follow a more direct trajectory where aposymbiotic (i.e., lacking symbionts) organisms are seeded with microbiota from other members within their population, such as by means of coprophagy in the case of triatomine bugs [11]. Alternatively, symbionts may be transmitted vertically. For instance, in the tsetse, the bacteria *Wolbachia* are transmitted through the germline, while the bacterium *Wigglesworthia* is transmitted via milk glands [12], an accessory organ that has been modified to provide nutrition to developing *in utero* larva. Finally, symbionts may use an integration of horizontal and vertical transmission, referred to as mixed-mode transmission, towards host establishment [13] where the acquisition of a more diverse symbiont population may be realized.

Symbioses are dynamic in nature in the sense that alterations in composition and function may occur during the host lifespan; with several obstacles needing to be overcome by potential symbionts to evolve as permanent partners of their host. Insect studies provide examples of changes in symbiont composition during the host lifespan. In *Drosophila* these changes have been associated with the intestine dysfunction in aging hosts [14], and in Lepidoptera, the microbiota has been tracked during development showing that not only the microbial composition but also its metabolic activities change across life stages [15]. For these associations to evolve, the bacteria have to successfully compete with other host-associated microorganisms, avoid or tolerate host immune system recognition, swap resources efficiently with the host, and develop a way to be transmitted to new hosts [16]. These partnerships may show adaptation towards partners through co-diversification, as hosts and symbionts have been subject to the same selective pressures, where collectively they may be referred to as a holobiont [2].

Genome structure of bacterial symbionts

Bacteria are prominent symbionts of animals and there are some particular features regularly present in the genomes of these symbionts that have co-evolved with their host for an evolutionary significant portion of time. First, there is the genome size reduction [17,18]. Usually, the genomes of obligate primary endosymbionts are smaller than their free-living relatives. The reduction process occurs when some genes that are not essential for survival within the host accumulate so many

mutations through relaxed selection that they become no longer functional (pseudogenes) and are eventually purged from the genome [19]. Second, the gene density, specifically, the fraction of the DNA that codes for proteins, is greater in these symbionts [20] due, at least in part, to their compact genomes. Third, in comparison to free-living bacteria, these symbionts typically show an adenine-thymine (AT) bias [21], within their genomes, which may result in an amino acid bias in the protein composition.

Genome interaction within the holobiont

A common feature of metabolism in insect symbioses [9] is that the biochemical pathways for essential metabolite synthesis are catalyzed by enzymes shared between host and symbiont genomes, as often each partner is unable to exclusively synthesize the metabolite from precursors. This metabolic complementarity between partners is a hallmark of endosymbiosis, likely driving its specificity and persistence, and exemplified through enhanced fitness of the coupling versus living solely [22]. For example, in the whitefly, *Bemisia tabaci*, the bacterial symbiont, *Portiera aleyrodidarum*, lacks genes involved in folate synthesis (i.e., *folE, B, K P C*, and *A*), but these compounds can be provided by another bacterial symbiont, *Hamiltonella defensa*. Furthermore, key steps in the synthesis of amino acids (i.e., serine and proline) are missing in the genomes of both these symbionts but may be complemented by enzymes provided in the whitefly genome [23]. Another example occurs in the symbiosis of the aphid, *Acyrthosiphon pisum*, and its bacterial symbiont, *Buchnera aphidicola*, where the bacterium provides the host with essential amino acids which are synthesized using intermediates supplied by the host [24–26]. Similarly, it has been demonstrated that the vitamin B_5 (pantothenate) synthesis pathway in *A. pisum* also relies on *B. aphidicola*. For B_5 synthesis, alanine and pantoate are converted into pantothenate through *A. pisum panC* activity. The aphid host is capable of synthesizing alanine from aspartate; however, it relies on *B. aphidicola* for the provisioning of pantoate which is reduced from 2-dehydropantoate through *ilvC* activity [27]. Metabolic complementarity is also predicted in the tripartite nested symbiosis of the mealybug-*Tremblaya-Moranella*, where the integration of enzymes from the three genomes is necessary for the synthesis of the semi-essential amino acid arginine and essential amino acids, such as tryptophan, phenylalanine, and isoleucine [28]. Lastly, host-symbiont metabolic collaboration is observed in the weevils (Curculionidae). These insects harbor the bacteria, Candidatus *Nardonella*, which provides the host with tyrosine. This amino acid is an essential component of the cuticle in exoskeletons, one of the most defining features in beetles. The *Nardonella* symbiont has the necessary enzymes for tyrosine synthesis, except for the tyrosine transaminase (*tyrB*), which catalyzes the final step, where 4-hydroxyphenylpyruvate is converted to tyrosine. Interestingly, the weevil genes *GOT1A* and *GOT2A* carry out this conversion and transcripts are enriched within bacteriomes, in convenient proximity to the symbiont [29].

Tsetse (Diptera: Glossinidae)

Tsetse flies are Dipterans belonging to the superfamily of exclusive blood-feeders, Hippoboscoidea. Tsetse are exclusively grouped in the family Glossinidae, within the monophyletic genus *Glossina*, and are divided into four groups: morsitans, fusca, palpalis and austeni [30], containing approximately 33 taxa. Tsetse flies are found only in sub-Saharan Africa. Depending on the species, these flies may be found in the woodlands (fusca), adjacent to the coast (austeni), in the savannahs (morsitans) or along riverine habitats (palpalis), which spans the coast and mangrove swamps during the rainy season [31].

Tsetse reproduce via adenotrophic viviparity, which involves one fertilized egg developing within the mother's uterus per gonotrophic cycle. While the larva is in the uterus, nourishment is provided through secretions (primarily composed of lipids and proteins) produced by modified accessory glands (known as milk glands). The *in utero* larval development takes from 7- 12 days depending on environmental conditions. Maternal investment is so high that by the time a tsetse larva is deposited, it often outweighs the mother. Additionally, due to the high energetic costs associated with adenotrophic viviparity, tsetse reproductive output is relatively low, ranging from 8-10 pupae/lifespan [32]. Larval development proceeds *in utero* until it reaches the third instar when it is subsequently deposited into the soil and enters pupation.

The adult flies emerge from the pupae after 30 days, and they usually mate 3-5 days after emergence. Following mating, it takes ~ 16 days for a female to lay her first larva [30]. This slow reproduction rate is compensated by a high adult survival rate. The adult fly lives for approximately three to four months.

Tsetse fly symbionts

A key feature in the evolution of eukaryotes has been the spatial and temporal partitioning of biochemical processes for the purpose of regulation. This partitioning reaches an even higher level of complexity with the presence of endosymbionts and the necessity to coordinate their metabolism with host biology. Tsetse flies possess a relatively simple microbiota, mainly constituted of four different microbial species.

Trypanosoma spp

Tsetse flies are the obligate vectors of Old-World Trypanosomes. Trypanosomes are single-celled protozoan parasites that have been associated with the tsetse for about 35 million years [33]. These organisms are the causative agent of human African trypanosomiasis (HAT), commonly referred to as sleeping sickness [34], which is a debilitating condition caused by the parasitic invasion of the central nervous system. The disease is endemic to 36 countries in sub-Saharan Africa. Trypanosoma infections (*T. b. brucei*) may also occur in other animals, causing a wasting disease known as nagana.

These parasites are transmitted through a tsetse bite. In the human blood, the metacyclic stage differentiates into the slender form and the cell-cycle arrested stumpy form. Antigenic variation is a key element in the evasion of the immune system by many parasites, which, in the case of *T. brucei*, there are over 2000 variant surface glycoproteins within the genome [35]. These proteins form a dense coat on the parasite and are differentially expressed to avoid recognition. Although both slender and stumpy forms are taken up during an infected bloodmeal, it is only the stumpy form that goes on to develop within tsetse. Trypanosomes establish the infection within the tsetse midgut, which then progresses to the proventriculus, the foregut, and, finally, the salivary glands [36]. Once within the salivary glands, trypanosomes may be transmitted to naive hosts to further propagate the infection.

Differences in the dynamics of the *Trypanosoma* parasite within the tsetse vector have been observed. For instance, regarding species, it was found that *G. pallidipes* are more likely to have trypanosome infection progress into the salivary glands than *G. morsitans*. For *G. morsitans*, an established infection in the midgut of males is more likely to progress to the salivary glands relative

to age-matched females [36]. Overall, the tsetse infection rates are low, only ~3.4% of flies that feed on infected hosts are capable to develop an infection and transmit the parasite [37].

Wigglesworthia glossinidia

Tsetse flies, of both sexes, feed exclusively on vertebrate blood. The blood, although rich in amino acids and iron, is particularly poor in B vitamins [38], which are essential for animals. The provisioning of multiple B vitamins by digestive tract symbionts has enabled the restricted feeding ecology of the tsetse fly. The most prevalent member of the tsetse microbiota, *Wigglesworthia glossinidia* [39,40], inhabits the cytosol of specialized tsetse cells known as bacteriocytes. The bacteriocytes collectively comprise the bacteriome organ, a ring-shaped mass that surrounds the anterior midgut. The tsetse-*Wigglesworthia* association dates back approximately 50-80 million years [40] and has been accompanied by drastic size reduction in the *Wigglesworthia* genome [41,42]. Despite the small size, the genome retains the potential to synthesize multiple B vitamins, namely, thiamine (B_1), riboflavin (B_2), nicotinamide (B_3), pantothenic acid (B_5), pyridoxine (B_6), and folate (B_9) [42,43]. To date, *Wigglesworthia* has not been cultured in synthetic media.

As obligate mutualists, tsetse rely on *Wigglesworthia* for the optimal performance of several physiological processes primarily attributed to nutrition and immunity. The bacteriome is enriched in fly gene transcripts that belong to the transmembrane category [44], which includes amino acid transporters and multivitamin transporters, thus facilitating nutrient exchange between partners. In turn, *Wigglesworthia* transcripts are enriched for the metabolism of cofactors and vitamins, supporting a complementary nutritional role between partners. Similar studies examining field-collected *Glossina pallidipes* bacteriomes also revealed a correlated enrichment in genes involved in metabolism and transport, between tsetse and *Wigglesworthia* [45].

Wigglesworthia provides other B vitamins to tsetse. For example, thiamine (B_1) provisioned by *Wigglesworthia* likely plays a role in *Sodalis* population dynamics and preventing antagonism within the tsetse holobiont [46]. Similarly, pyridoxine from *Wigglesworthia* is necessary for fecundity, given that this vitamin is essential for proline homeostasis. Proline is the main energy source of the fly and is essential during the energetically demanding phase of lactation [47]. Further, the *Wigglesworthia* symbionts are also necessary for tsetse sexual maturity and optimal larval development due to folate (B_9) provisioning within *G. morsitans* [48]. Lastly, the *Wigglesworthia* association is also essential for the functioning of the tsetse immune system [49,50], given that when this symbiont is absent, melanization is compromised, hemocyte count is lower and there is reduced expression of genes involved in antimicrobial activity [51].

Sodalis glossinidius

S. glossinidius, like *Wigglesworthia,* is also a Gammaproteobacterium. This symbiont recently transitioned from a free-living bacterium to a secondary symbiont, as evidenced by its ability to be cultivated in pure culture [52] and its relatively larger genome (~4.2 Mb), albeit containing a significant number of pseudogenes [53,54]. *Sodalis* related symbionts have been identified in other orders besides Diptera, such as Hemiptera [55], Phtiraptera [56], and Coleoptera [57]. The *Sodalis* infections are located intra- and extra-cellularly in the midgut and other tissues of the tsetse such as muscle, fat body, milk glands, and salivary glands [58]. The prevalence of *Sodalis* in field flies may range, for instance, from 9.3% to 85% depending on the tsetse species [59] and the collection location. The *S. glossinidius* symbiont is thought to promote trypanosome (*Trypanosoma spp*.)

infection through midgut chitinase activity, resulting in the production of N-acetyl glucosamine (GlcNac). GlcNac binds to an anti-trypanosomal gut lectin generated by the fly hampering its antimicrobial role [60]. However, the presence of *S. glossinidius* is dispensable for *Trypanosoma* infection, given that flies negative for the bacterium can still harbor parasites. Correlation studies show contradictory outcomes regarding the presence of both *S. glossinidius* and trypanosomes in the field [59,61,62], ranging from no correlation, to correlation dependent on *Trypanosoma* species [63] or *S. glossinidius* genotype [64]. Removal of this bacterium via antibiotic treatment seems to have a detrimental effect on the longevity of the fly; however, an intrinsic effect of the antibiotic streptozotocin on the tsetse cannot be ruled out [60]. *Sodalis* does not have any known provisioning role in tsetse.

Wolbachia pipientis

Wolbachia is an Alphaproteobacteria which is a reproductive parasite that causes cytoplasmic incompatibility (CI) in tsetse [65]. This phenomenon enables reproductive success when a *Wolbachia*-infected female mates with an uninfected male or with a male infected with the same *Wolbachia* strain. However, when only the male is infected or when mated females and males are infected with different strains, the embryos die. Therefore, CI confers a reproductive advantage to infected females, as the bacteria are vertically transmitted by females [66]. In the field, and similar to *Sodalis* field distributions, tsetse flies show a variable prevalence for *Wolbachia* infections, ranging from 0% to 100% depending on the fly species [66]. In *G. morsitans* and *G. brevipalpis*, *Wolbachia* is confined to reproductive tissues, although it has been found in the somatic tissues of *G. austeni* [67]. Horizontal gene transfer events from the *W. pipientis* genome to the *G. morsitans* X, Y, and B chromosomes have been reported [68].

DNA methylation due to symbiosis

DNA methylation is an epigenetic modification by which the 5' position of cytosine in the genomic DNA is methylated [69,70]. Methylation typically occurs within CpG islands, although, other sequence contexts may also demonstrate this epigenetic alteration[71,72]. Methylation occurs within promoter and enhancer regions of vertebrates, thereby downregulating the initiation of transcription [73]. However, in insects, methylation preferences regarding genomic location vary [74]. Insect physiological processes where DNA methylation has been implicated include the caste determination of honeybees, sexual dimorphism in planthoppers, gene regulation in termites, and especially interesting in the context of co-evolution and vector biology, towards female fecundity, embryo development, and immune response [75]. Although, it has been established that DNA methylation plays crucial biological roles within several insect orders, a comprehensive understanding of the determinants of DNA methylation patterns lacking [76]. In Diptera, there are contradictory results regarding the methylation status of genomes. For example, there are empirical studies confirming methylation in *Drosophila melanogaster* [42,72]. Yet, the absence of canonical *de novo* and maintenance DNA methyltransferases (i.e., DNMT3 and DNMT1; respectively), reject the presence of DNA methylation within the order [75,77–79]. However, DNA methylation has been found in *Bombyx mori* (Lepidoptera), where no *de novo* DNA-methyltransferase (i.e., DNMT3) is predicted from the genome [74]. It is worth noting that the methylation reported within the *D. melanogaster* genome occurs in a pattern unlike that seen in other eukaryotes, being the highest 2-3 hours post-fertilization [80], with highly methylated genes significantly enriched for gene ontology (GO) terms associated with morphological development and transcription factors [72]. This demonstrates a gap in our current understanding of both the extent and the mechanisms by which

DNA-methylation in other taxa may occur. Particularly, where canonical *de novo* DNA-methyltransferases are absent, and potentially where cryptic DNA methyltransferases may exist, such as in *Glossina* [75,81].

Folate (B_9) is critical for 1-carbon metabolism which provides the source of methyl groups for methylation reactions, including DNA methylation. More specifically, 5-methyl tetrahydrofolate transfers its methyl group to homocysteine to produce methionine. Methionine is then subsequently transformed to the methyl donor S-adenosyl-methionine (SAM), which then supplies the methyl group that will be donated to cytosine through catalysis of the DNA-methyltransferase and the production of S-adenosyl homocysteine [82]. Because of its established role in folate provisioning [48], *Wigglesworthia* may be hypothesized to affect host genome methylation status.

Next-generation sequencing has greatly enabled characterizing the microbiota relative to both composition and functional roles. For instance, RNA-Seq analyses have facilitated the identification of differentially regulated genes in host conditions impacted by symbionts, such as through the comparison of aposymbiotic (i.e., devoid of symbionts) versus symbiotic hosts, and even within tissues that are symbiont-enriched, such as the host bacteriome. Once differentially expressed genes have been identified, more targeted functional analyses, for example, fine scale temporal-spatial profiling and disruption assays (i.e., mutagenesis, chemical disruptions, etc.) can be performed. To date, little is known concerning how symbiont-host activity is regulated, particularly the effect of the microbiota towards DNA methylation status of the host genome.

Symbiosis in a changing environment

Environmental change is a major driver of evolution. Initially, increased sunlight facilitated the expansion of eukaryotes with photosynthetic plastids, while increased oxygen opened new niches for eukaryotes associated with proto-mitochondria. Subsequently, as the abundance of vascular plants and animals with circulatory systems grew, symbioses allowed insects to exploit specialized and nutritionally imbalanced food sources, such as plant sap or blood. Adaptation via partnership between two very different lineages, such as within insects and bacteria, facilitates a faster response to environmental changes. For example, insecticide use in sugarcane field favored the alliance between the bean bug, *Riptortus pedestris*, and a fenitrothion-degrading *Burkholderia* strain where the insecticide is rendered harmless as it is assimilated by the bacterium as a carbon source [83]. Microbial symbiosis allows partners to take advantage of preexisting mechanisms to use for their own survival.

Conclusions

To understand the full range of insect-microbiota interactions, it is necessary to consider the genetic potential of all partners involved in symbiosis. The biology and natural symbiotic associations of the tsetse fly provide an ideal proxy for the comparative study of the evolution of symbioses. Similarly, this system allows for empirical studies to dissect the regulatory mechanisms that coordinate several aspects of insect-microbiota biology, such as nutrient exchange and permissiveness to parasites. Besides providing the basis for a deeper understanding of ecological and organismal biology features, the study of symbioses, particularly in blood-feeding vectors is of great consequence for epidemiological studies and the design of vector control strategies aiming at halting transmission of vector-borne diseases.

References

1. Frank, B. Über die auf Wurzelsymbiose beruhende Ernährung gewisser Bäume durch unterirdische Pilze from Berichte der Deutschen Botanischen Gesellschaft. *Mycorrhiza* **3**, 128–145 (1885).
2. Marchesi, J. R. & Ravel, J. The vocabulary of microbiome research: a proposal. *Microbiome* **3**, 31 (2015).
3. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998).
4. Lederberg, B. J. & McCray, A. T. ' Ome Sweet ' Omics-- A Genealogical Treasury of Words. *Sci.* **15**, 8 (2001).
5. Ursell, L. K., Metcalf, J. L., Parfrey, L. W. & Knight, R. Defining the human microbiome. *Nutr. Rev.* **70**, (2012).
6. Whiteside, S. A., Razvi, H., Dave, S., Reid, G. & Burton, J. P. The microbiome of the urinary tract—a role beyond infection. *Nat. Rev. Urol.* **12**, 81–90 (2015).
7. Su, Q., Zhou, X. & Zhang, Y. Symbiont-mediated functions in insect hosts. *Commun. Integr. Biol.* **6**, 1–7 (2013).
8. Jiggins, F. M. & Hurst, G. D. D. Rapid Insect Evolution by Symbiont Transfer. *Science (80-.).* **332**, 185–186 (2011).
9. Wilson, A. C. C. & Duncan, R. P. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc. Natl. Acad. Sci.* **112**, 10255–10261 (2015).
10. Moran, N. a & Telang, a. Bacteriocyte-associated symbionts of insects - A variety of insect groups harbor ancient prokaryotic endosymbionts. *Bioscience* **48**, 295–304 (1998).
11. Matthews, S., Rao, V. S. & Durvasula, R. V. Modeling horizontal gene transfer (HGT) in the gut of the Chagas disease vector Rhodnius prolixus. *Parasit. Vectors* **4**, 77 (2011).
12. Attardo, G. M. *et al.* Analysis of milk gland structure and function in Glossina morsitans: Milk protein production, symbiont populations and fecundity. *J. Insect Physiol.* **54**, 1236–1242 (2008).
13. Ott, B. M., Dacks, A. M., Ryan, K. J. & Rio, R. V. M. A tale of transmission: Aeromonas veronii activity within leech-exuded mucus. *Appl. Environ. Microbiol.* **82**, 2644–2655 (2016).
14. Clark, R. I. *et al.* Distinct Shifts in Microbiota Composition during Drosophila Aging Impair Intestinal Function and Drive Mortality. *Cell Rep.* **12**, 1656–1667 (2015).
15. Chen, B. *et al.* Biodiversity and Activity of the Gut Microbiota across the Life History of the Insect Herbivore Spodoptera littoralis. *Sci. Rep.* **6**, 29505 (2016).
16. Sachs, J. L., Skophammer, R. G. & Regus, J. U. Evolutionary transitions in bacterial symbiosis. *Proc. Natl. Acad. Sci.* **108**, 10800–10807 (2011).
17. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011).
18. Bennett, G. M., McCutcheon, J. P., MacDonald, B. R., Romanovicz, D. & Moran, N. A. Differential genome evolution between companion symbionts in an Insect-Bacterial symbiosis. *MBio* **5**, 1–12 (2014).
19. Burke, G. R. & Moran, N. A. Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. *Genome Biol. Evol.* **3**, 195–208 (2011).
20. Nakabachi, A. *et al.* Bacterial Endosymbiont Carsonella. *Science* **314**, 2006–2008 (2008).

21. Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and Evolution of Heritable Bacterial Symbionts. *Annu. Rev. Genet.* **42**, 165–190 (2008).
22. Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T. & West, S. A. The evolution of host-symbiont dependence. *Nat. Commun.* **8**, 15973 (2017).
23. Rao, Q. *et al.* Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly Bemisia tabaci. ??? 1–13 (2015). doi:10.1186/s12864-015-1379-6
24. Shigenobu, S. & Wilson, A. C. C. Genomic revelations of a mutualism: The pea aphid and its obligate bacterial symbiont. *Cell. Mol. Life Sci.* **68**, 1297–1309 (2011).
25. Wilson, A. C. C. *et al.* Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola. *Insect Mol. Biol.* **19**, 249–258 (2010).
26. Russell, C. W., Bouvaine, S., Newell, P. D. & Douglasa, A. E. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Appl. Environ. Microbiol.* **79**, 6117–6123 (2013).
27. Price, D. R. & Wilson, A. C. A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol.* **12**, 110 (2014).
28. McCutcheon, J. P. & Von Dohlen, C. D. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol.* **21**, 1366–1372 (2011).
29. Anbutsu, H. *et al.* Small genome symbiont underlies cuticle hardness in beetles. *Proc. Natl. Acad. Sci.* **114**, E8382–E8391 (2017).
30. Krafsur, E. S. Tsetse flies: Genetics, evolution, and role as vectors. *Infect. Genet. Evol.* **9**, 124–141 (2009).
31. Solano, P., Ravel, S. & de Meeûs, T. How can tsetse population genetics contribute to African trypanosomiasis control? *Trends Parasitol.* **26**, 255–263 (2010).
32. Benoit, J. B., Attardo, G. M., Baumann, A. A., Michalkova, V. & Aksoy, S. Adenotrophic Viviparity in Tsetse Flies: Potential for Population Control and as an Insect Model for Lactation. *Annu. Rev. Entomol.* **60**, 351–371 (2015).
33. Steverding, D. The history of African trypanosomiasis. *Parasit. Vectors* **1**, 3 (2008).
34. Castellani, A. On the discovery of a species of trypanosoma in the cerebrospinal fluid of cases of sleeping sickness. *Lancet* **161**, 1735–1736 (1903).
35. Matthews, K. R., McCulloch, R. & Morrison, L. J. The within-host dynamics of African trypanosome infections. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140288 (2015).
36. Peacock, L., Ferris, V., Bailey, M. & Gibson, W. The influence of sex and fly species on the development of trypanosomes in tsetse flies. *PLoS Negl. Trop. Dis.* **6**, (2012).
37. Simwango, M. *et al.* Molecular prevalence of trypanosome infections in cattle and tsetse flies in the Maasai Steppe, northern Tanzania. *Parasit. Vectors* **10**, 507 (2017).
38. Douglas, A. E. The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. *Curr. Opin. Insect Sci.* **23**, 65–69 (2017).
39. Aksoy, S. Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *Int. J. Syst. Evol. Microbiol.* **45**, 848–851 (1995).
40. Chen, X., Li, S. & Aksoy, S. Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *J. Mol. Evol.* **48**, 49–58 (1999).
41. Rio, R. V. M. *et al.* Insight into the Transmission Biology and Species-Specific Functional

Capabilities of Tsetse (Diptera: Glossinidae) Obligate Symbiont Wigglesworthia. *MBio* **3**, 1–13 (2012).
42. Akman, L. *et al.* Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat. Genet.* **32**, 402–407 (2002).
43. Rio, R. V. M., Symula, R. E. & Wang, J. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont Wigglesworthia. *MBio* **3**, e00240-11 (2012).
44. Bing, X. *et al.* Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia : transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proc. R. Soc. B Biol. Sci.* **284**, 20170360 (2017).
45. Medina Munoz, M., Pollio, A. R., White, H. L. & Rio, R. V. M. Into the Wild: Parallel Transcriptomics of the Tsetse-Wigglesworthia Mutualism within Kenyan Populations. *Genome Biol. Evol.* **9**, 2276–2291 (2017).
46. Snyder, A. K., Deberry, J. W., Runyen-Janecky, L. & Rio, R. V. M. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. R. Soc. B Biol. Sci.* **277**, 2389–2397 (2010).
47. Michalkova, V., Benoit, J. B., Weiss, B. L., Attardo, G. M. & Aksoy, S. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Appl. Environ. Microbiol.* **80**, 5844–5853 (2014).
48. Snyder, A. K. & Rio, R. V. M. 'Wigglesworthia morsitans' folate (vitamin B<inf>9</inf>) biosynthesis contributes to tsetse host fitness. *Appl. Environ. Microbiol.* **81**, 5375–5386 (2015).
49. Wang, J., Wu, Y., Yang, G. & Aksoy, S. Interactions between mutualist Wigglesworthia and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proc. Natl. Acad. Sci.* **106**, 12133–12138 (2009).
50. Weiss, B. L., Wang, J., Maltz, M. A., Wu, Y. & Aksoy, S. Trypanosome Infection Establishment in the Tsetse Fly Gut Is Influenced by Microbiome-Regulated Host Immune Barriers. *PLoS Pathog.* **9**, e1003318 (2013).
51. Weiss, B. L., Wang, J. & Aksoy, S. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol.* **9**, e1000619 (2011).
52. 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. Evol. Microbiol.* **49**, 267–275 (1999).
53. Toh, H. *et al.* Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome Res.* **16**, 149–156 (2006).
54. Belda, E., Silva, F. J., Peretó, J. & Moya, A. Metabolic networks of Sodalis glossinidius: A systems biology approach to reductive evolution. *PLoS One* **7**, (2012).
55. Hosokawa, T., Kaiwa, N., Matsuura, Y., Kikuchi, Y. & Fukatsu, T. Infection prevalence of Sodalis symbionts among stinkbugs. *Zool. Lett.* **1**, 5 (2015).
56. Boyd, B. M. *et al.* Louse Proechinophthirus fluctus (Phthiraptera : Anoplura). **82**, 3185–3197 (2016).
57. Snyder, A. K., Mcmillen, C. M., Wallenhorst, P. & Rio, R. V. M. The phylogeny of Sodalis-like symbionts as reconstructed using surface-encoding loci. *FEMS Microbiol. Lett.* **317**, 143–151 (2011).
58. Balmand, S., Lohs, C., Aksoy, S. & Heddi, A. Tissue distribution and transmission routes

for the tsetse fly endosymbionts. *J. Invertebr. Pathol.* **112**, 116–122 (2013).
59. Dennis, J. W. *et al.* Sodalis glossinidius prevalence and trypanosome presence in tsetse from Luambe National Park, Zambia. *Parasit. Vectors* **7**, 378 (2014).
60. Dale, C. & Welburn, S. C. The endosymbionts of tsetse flies: manipulating host-parasite interactions. *Int J Parasitol* **31**, 628–631 (2001).
61. Geiger, A., Ravel, S., Frutos, R. & Cuny, G. Sodalis glossinidius (Enterobacteriaceae) and vectorial competence of Glossina palpalis gambiensis and Glossina morsitans morsitans for Trypanosoma congolense Savannah type. *Curr. Microbiol.* **51**, 35–40 (2005).
62. Farikou, O. *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.* **10**, 115–121 (2010).
63. Wamwiri, F. N. *et al.* Infection with the secondary tsetse-endosymbiont Sodalis glossinidius (Enterobacteriales: Enterobacteriaceae) influences parasitism in Glossina pallidipes (Diptera: Glossinidae). *J. Insect Sci.* **14**, 14–17 (2014).
64. Geiger, A. *et al.* Vector competence of Glossina palpalis gambiensis for Trypanosoma brucei s.l. and genetic diversity of the symbiont Sodalis glossinidius. *Mol. Biol. Evol.* **24**, 102–109 (2007).
65. Alam, U. *et al.* Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the Tsetse fly glossina morsitans. *PLoS Pathog.* **7**, (2011).
66. Doudoumis, V. *et al.* Tsetse-Wolbachia symbiosis: Comes of age and has great potential for pest and disease control. *J. Invertebr. Pathol.* **112**, S94–S103 (2013).
67. Cheng Q et al. Tissue distribution and prevalence of Wolbachia infections in tsetse flies, Glossina spp.pdf. *Med Vet Entomol.* **14**, 44–50 (2000).
68. Brelsfoard, C. *et al.* Presence of Extensive Wolbachia Symbiont Insertions Discovered in the Genome of Its Host Glossina morsitans morsitans. *PLoS Negl. Trop. Dis.* **8**, (2014).
69. McGhee, J. & Ginder, G. McGhee - Specific DNA methylation sites in the vicinity of the chicken beta-globin genes.pdf. *Nature* **280**, 419–420 (1979).
70. Taylor, S. M. & Jones, P. a. Cellular differentiation. *Int. J. Obes.* **9 Suppl 1**, 15–21 (1985).
71. Kunert, N., Marhold, J., Stanke, J., Stach, D. & Lyko, F. A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development* **130**, 5083–5090 (2003).
72. Takayama, S. *et al.* Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. *Genome Res.* **24**, 821–830 (2014).
73. Bird, A. P. & Wolffe, A. P. Methylation-induced repression-belts, braces, and chromatin. *Cell* **99**, 451–454 (1999).
74. Glastad, K. M., Hunt, B. G. & Goodisman, M. A. Evolutionary insights into DNA methylation in insects. *Curr. Opin. Insect Sci.* **1**, 25–30 (2014).
75. Bewick, A. J., Vogel, K. J., Moore, A. J. & Schmitz, R. J. Evolution of DNA methylation across insects. *Mol. Biol. Evol.* **34**, 654–665 (2017).
76. Kim, D., Thairu, M. W. & Hansen, A. K. Novel Insights into Insect-Microbe Interactions—Role of Epigenomics and Small RNAs. *Front. Plant Sci.* **7**, 1–18 (2016).
77. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–9 (2010).
78. Raddatz, G. *et al.* Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci.* **110**, 8627–8631 (2013).
79. Glastad, K. M., Hunt, B. G., Yi, S. V. & Goodisman, M. A. D. DNA methylation in insects: On the brink of the epigenomic era. *Insect Mol. Biol.* **20**, 553–565 (2011).

80. Lyko, F., Ramsahoye, B. H. & Jaenisch, R. DNA methylation in Drosophila melanogaster. *Nature* **408**, 538–540 (2000).
81. Watanabe, J. *et al.* Genome sequence of the tsetse fly (Glossina morsitans): Vector of African trypanosomiasis. *Science (80-.).* **344**, 380–386 (2014).
82. Tserga, A., Binder, A. M. & Michels, K. B. Impact of folic acid intake during pregnancy on genomic imprinting of *IGF2/H19* and 1-carbon metabolism. *FASEB J.* fj.201601214RR (2017). doi:10.1096/fj.201601214RR
83. Tago, K., Okubo, T., Itoh, H., Kikuchi, Y., Hori, T., Sato, Y., Nagayama, A., Hayashi, K., Ikeda, S., & Hayatsu, M. Insecticide-degrading Burkholderia symbionts of the stinkbug naturally occupy various environments of sugarcane fields in a Southeast island of Japan. *Microbes and environments*, *30*(1), 29–36. (2015). https://doi.org/10.1264/jsme2.ME14124

CHAPTER 2: Quorum sensing sets the stage for the establishment and vertical transmission of *Sodalis praecaptivus* in tsetse flies

An original research article published in *PLOS Genetics* from PLOS: 16(8): e1008992

Miguel Medina Munoz[1], Noah Spencer[1], Shinichiro Enomoto[2], Colin Dale[2] and Rita V.M. Rio[1,#]
[1]Department of Biology, Eberly College of Arts and Sciences, West Virginia University, Morgantown, WV, 26505, USA
[2]Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112, USA

Abstract

Bacterial virulence factors facilitate host colonization and set the stage for the evolution of parasitic and mutualistic interactions. The *Sodalis*-allied clade of bacteria exhibit striking diversity in the range of both plant and animal feeding insects they inhabit, suggesting the appropriation of universal molecular mechanisms that facilitate establishment. Here, we report on the infection of the tsetse fly by free-living *Sodalis praecaptivus*, a close relative of many *Sodalis*-allied symbionts. Key genes involved in quorum sensing, including the homoserine lactone synthase (*ypeI*) and response regulators (*yenR* and *ypeR*) are integral for the benign colonization of *S. praecaptivus*. Mutants lacking *ypeI*, *yenR*, and *ypeR* compromised tsetse survival as a consequence of their inability to repress virulence. Genes under regulation via quorum sensing, including homologs of the binary insecticidal toxin PirAB and a putative symbiosis-promoting factor CpmAJ, demonstrated negative and positive impacts, respectively, on tsetse survival. Taken together with results obtained from experiments involving weevils, this work shows that quorum sensing virulence suppression plays an integral role in facilitating the establishment of *Sodalis*-allied symbionts in diverse insect hosts. This knowledge contributes to the understanding of the early evolutionary steps involved in the formation of insect-bacterial symbiosis. Further, despite having no established history of interaction with tsetse, *S. praecaptivus* can infect reproductive tissues, enabling vertical transmission through adenotrophic viviparity within a single host generation. This creates an option for the use of *S. praecaptivus* in the biocontrol of insect disease vectors via paratransgenesis.

Author summary

Symbiosis drives organismal novelty. Yet, we know little about the origin, establishment, and persistence of symbiosis between animals and bacteria. *Sodalis*-allied symbionts have established independent infections in many insects. Here we show that quorum sensing facilitates the establishment of a novel *Sodalis praecaptivus* infection in tsetse flies, in a fashion very similar to that observed in weevils. Importantly, quorum sensing modulation of virulence allows *Sodalis* to establish in these hosts with minimal pathology and may explain the propensity for these symbionts to adopt associations with a wide range of hosts. Furthermore, *S. praecaptivus* infects reproductive tissues, enabling vertical transmission within a single host generation, potentially facilitating the use of *S. praecaptivus* in the control of insect disease vectors via paratransgenesis.

Introduction

Bacteria occupy countless niches, including numerous adaptive (mutualistic) associations with plants and animals. Here, these microbes play numerous beneficial roles in host biology, impacting development [1, 2], defense [3], and enhancing immunity and nutrition ([4], reviewed in [5]). Apart from a few examples provided by symbioses that rely on environment-mediated transmission, such as the *Vibrio*-squid and the *Rhizobium*-legume associations [6, 7], molecular mechanisms that enable the initial establishment and subsequent vertical transmission remain largely obscure [8]. Bacterial genera that are capable of infecting a wide range of arthropods and persisting by vertical transmission (such as *Wolbachia*, *Arsenophonus*, *Rickettsia*, and *Sodalis* [9-12]) offer insight into these molecular features. Deciphering these key traits that facilitate host relations is useful for developing applications which require the establishment and maintenance of genetically modified symbionts within a host, such as the use of probiotics to restore the composition of gut flora or the implementation of paratransgenesis to express novel genes in a host using genetically engineered symbionts [13, 14].

Closely related *Sodalis*-allied symbionts have been identified within numerous insect orders. A spectrum of *Sodalis*-insect interactions ranging from enigmatic and facultative to mutualistic and obligate has arisen through multiple independent infection events [9, 15-24]. Unlike other symbiont clades in which close non insect-associated members of the genus are not known, the genus *Sodalis* includes *S. praecaptivus*, which is viewed and has been experimentally adopted as a closely-related environmental antecedent to the *Sodalis*-allied symbionts found in many insect taxa [25]. *S. praecaptivus* was cultured fortuitously from a human hand infection following impalement with a dead tree branch [25]. This bacterium was identified using 16S rRNA analysis that revealed high sequence identity (99%) to the recently derived endosymbionts of various weevil and stink bug species [25]. The identification of an environmental reservoir for *Sodalis*-allied bacteria further supports a source-sink model [26] for the transfer of a free-living bacterium (i.e. source) into diverse insect hosts (i.e. sinks) in which mutualistic associations can then evolve. Importantly, the capacity to maintain *S. praecaptivus* in culture coupled with its amenability towards genetic manipulation provides an ideal model to further our understanding of key factors enabling the commencement and progression of symbioses in a wide range of insects.

The *S. praecaptivus* genome [25] is considerably larger than other *Sodalis*-allied symbionts and contains substantially more intact homologs of virulence and toxin genes that are typically associated with animal and plant pathogens. The reduced genomes of the *Sodalis*-allied endosymbionts are subsets of *S.* praecaptivus, derived following transition to restricted insect-associated lifestyles. Thus, the larger gene inventory found in *S. praecaptivus* is thought to allow greater environmental and host plasticity, some of which may facilitate host-specific exchanges instrumental towards the establishment of insect symbioses. Interestingly, *S. praecaptivus* is able to colonize within grain weevils through the use of quorum sensing that limits the expression of virulent insecticidal genes to only within the incipient stages of infection enabling population growth and persistence by gaining access to host tissues and cells [27].

Quorum sensing is a form of intercellular communication that enables bacteria to determine their local population density, and coordinate collective group behavior such as swarming, motility, and light production, through the simultaneous regulation of gene expression across the population [28,

[29]. Quorum sensing is directed by the synthesis and subsequent diffusion of small molecules known as autoinducers (e.g. N-acyl homoserine lactone signaling molecules for some Gram-negative bacteria) that can diffuse in and out of the bacterial cell and modulate transcription of target genes via interaction with specific protein response regulators [30]. The disruption of quorum sensing within *S. praecaptivus* results in a rapid and potent host killing phenotype upon microinjection into weevils [27]. This occurs because quorum sensing normally represses the expression of virulence factors associated with insect killing, including the PirAB insecticidal toxins, chitinases, and collagenase-like proteins [27]. Under natural circumstances, these virulence factors are expressed only at low bacterial population density (i.e., during the initial stage of infection), presumably to allow bacteria to gain access into host tissues and cells [27].

While the results obtained from studies involving weevils are intriguing, they currently represent an isolated case and the significance of the quorum sensing control of virulence towards the colonization of *S. praecaptivus* in a broader range of insects has not been explored. Here, we examine the ability of *S. praecaptivus* to establish in a member of a distinct order of insects, the Dipteran tsetse fly *Glossina morsitans*, which also harbors an autochthonous *Sodalis*-allied symbiont (*S. glossinidius*). The colonization, persistence, and localization of *S. praecaptivus* within tsetse flies were evaluated relative to bacterial population density and host life history traits, specifically lifespan and fecundity. We find that key genes in quorum sensing, including those encoding a homoserine lactone synthase (*ypeI*) and response regulators (*yenR* and *ypeR*) also play an integral role in virulence restraint in tsetse, with *S. praecaptivus* mutants (lacking these genes) producing significantly greater tsetse mortality following their introduction. These results are consistent with the notion that quorum sensing functions to control the virulence of *Sodalis* in a broad range of insects. Our work also shows that, despite having no established history of interaction with tsetse, *S. praecaptivus* infects tsetse reproductive tissue, enabling vertical transmission within a single host generation. The amenability of *S. praecaptivus* towards genetic engineering [27], its ease of establishment within tsetse coupled with its ability to be vertically transmitted (albeit currently at low levels), provides a new option for the development and implementation of paratransgenic tools for disease control, perhaps including tsetse transmitted trypanosomiasis. In addition to enhancing our understanding of molecular features involved in the initiation of symbiosis, *S. praecaptivus* also provides a model for future studies of the mechanistic basis of vertical transmission, which is poorly understood in insects.

Results

S. praecaptivus establishes a persistent and benign infection within tsetse

Successful establishment of a microbe in a host is dependent on interactions with immunological responses, nutritional resources, and resident microbiota [23, 26, 31]. It was previously shown that *S. praecaptivus*, a free-living relative of the *Sodalis*-host associated symbionts found within many different insects, was capable of sustaining infections within *Si. zeamais* grain weevils cleared of their endogenous *Ca.* S. pierantonius symbionts [27]. Here, we subjected tsetse flies to a similar infection regimen to determine if they could also be infected with *S. praecaptivus*. In this case, *S. praecaptivus* was introduced into flies either *per os* in heat-inactivated (HI) blood or by intrathoracic microinjection at concentrations similar to those previously used for the inoculation of foreign bacteria into tsetse [27, 32] (Table 1).

Table 1. *S. praecaptivus* infection of tsetse.
The percentage of tsetse infected with WT *S. praecaptivus* following either introduction through *per os* supplementation of heat-inactivated blood meal or thoracic microinjection. *G. morsitans*^{STZ} = *S. glossinidius*-free *G. morsitans*. Fly infection status was determined 7 days following WT (CD14) *S. praecaptivus* introduction.

Mode of introduction	CFU introduced per fly/ tsetse line	% of flies infected (confirmed infections/flies challenged)
per os	1x10^3 /*G. morsitans*	79% (107/135)
	1x10^5 /*G. morsitans*	87% (39/45)
	1x10^3 /*G. morsitans*^{STZ}	75% (9/12)
	1x10^5 / *G. morsitans*^{STZ}	100% (12/12)
microinjection	1x 10^5 /*G. morsitans*	100% (63/63)
	1x10^5 / *G. morsitans*^{STZ}	89% (8/9)

https://doi.org/10.1371/journal.pgen.1008992.t001

Tsetse flies were examined 7 d following bacterial challenge for the presence of infection by plating surface-sterilized homogenized flies and PCR amplifying the DNA of resulting bacterial colonies with *S. praecaptivus*-specific trans-aconitate 2-methyltransferase (*tam*) gene primers. Tsetse flies were found to be capable of sustaining *S. praecaptivus* infections throughout our initial (7d) observation window using both *per os* (~80% of tsetse individuals receiving a 1 x 10^3 CFU introduction of WT *S. praecaptivus* harbored infections while a slightly higher infection rate of 87% was achieved when challenging tsetse individuals with a 1 x 10^5 CFU introduction of WT *S. praecaptivus*, Table 1) and intrathoracic microinjection (100% of challenged tsetse individuals were infected at 7 d with introduction of 1 x 10^5 CFU of WT *S. praecaptivus*, Table 1). Microinjection into tsetse hemolymph proved slightly more efficacious towards *S. praecaptivus* establishment supported by the higher prevalence of infection among individuals in comparison to oral introduction (100% versus 87% of individuals at a 1 x 10^5 CFU delivery, respectively, Table 1). The microinjection of *S. praecaptivus* also resulted in significantly greater CFU per fly in comparison to *per os* at 7 d post introduction (Fig. 1A, Mann-Whitney U test, $p = 0.004$). Furthermore, infections were found to persist throughout the tsetse lifespan as *S. praecaptivus* presence was observed at 50 d post microinjection (Fig. 1B). The *S. praecaptivus* infections also appear to be benign with respect to host lifespan, as no adverse effects towards tsetse survival were observed when compared to noninjected (Log-rank test, $p = 0.2250$) and mock-injected controls (Log-rank test, $p = 0.1526$) (Fig. 1C).

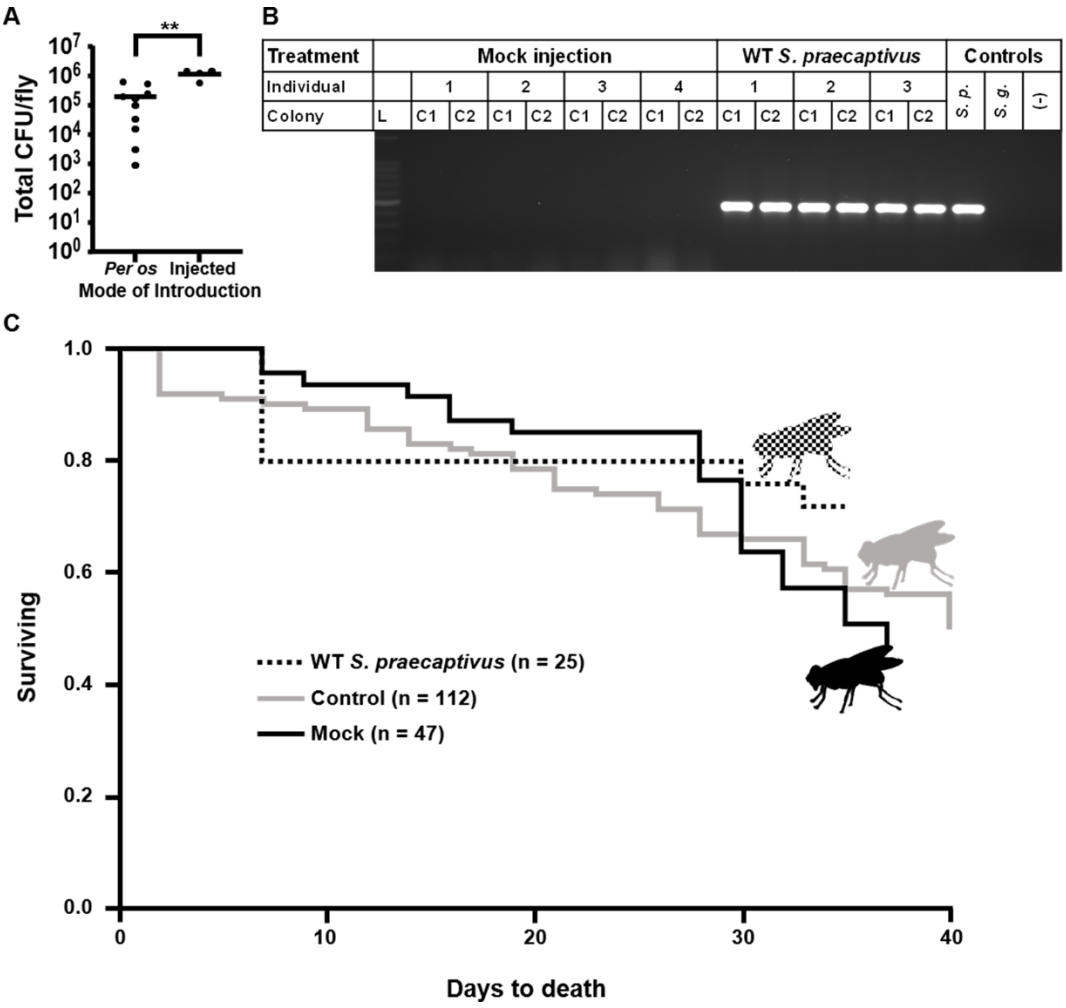

Figure 1. *S. praecaptivus* can establish long lasting tsetse infections.
A. Total WT *S. praecaptivus* density per fly is greater when individuals are challenged by microinjection versus *per os*. Both groups were administered 1 x 10^5 bacteria. The density of WT *S. praecaptivus* flies following 7 d post introduction was assessed via CFU counts after surface sterilization, homogenization, serial dilution and plating. Horizontal lines represent mean bacterial count per treatment, while dots indicate the bacterial density within individual flies. The Y axis (Total CFU/fly) is plotted in logarithmic scale. Mean bacterial density between introduction modes was compared through a Mann-Whitney U test (**, p = 0.004). B. PCR for the *S. praecaptivus* specific *tam* gene (Genbank: AHF76984.1, 442 bp amplicon) confirmed that colonies from plating of WT *S. praecaptivus*-injected fly homogenates 50 days post microinjection were *S. praecaptivus*, indicating persistent infections. Sampled colonies from mock-injected fly homogenates showed no evidence of *S. praecaptivus* infection. Individual = distinct flies; Colony = two bacterial colonies per fly were randomly selected for PCR verification; L = DNA molecular marker; *S.p.* = *S. praecaptivus* DNA; *S.g.* = *S. glossinidius* DNA; (-) = negative control corresponding to a no DNA template reaction. This assay was carried out for > 10 individuals per group, and one representative analysis is shown here. C. The established WT *S. praecaptivus* infection does not impact tsetse survival. Kaplan-Meier curves comparing survival of tsetse injected with *S. praecaptivus* to those of noninjected and mock injected control lines. During the observation period there were no significant differences between the survival of control flies and those microinjected with WT *S. praecaptivus* (*S. praecaptivus* vs. mock injected; Log-rank test, p = 0.1526, *S. praecaptivus* vs. control [non-injected flies]; Log-rank test, p = 0.2250). n = flies per treatment group. WT *S. praecaptivus* used was CD14 (Table 2, Appendix A).
https://doi.org/10.1371/journal.pgen.1008992.g001

Quorum sensing is essential for *S. praecaptivus* persistence within tsetse

In *S. praecaptivus*, the homoserine lactone (3-oxo-hexanoyl homoserine lactone; OHHL) is synthesized by a homolog of the canonical LuxI synthase, designated YpeI, where it is sensed by two response regulator homologs, YpeR and YenR [27]. When OHHL density surpasses a threshold, *ypeR* and *yenR* modulate gene expression profiles at the population level, downregulating the expression of virulent toxins and preventing host killing [27]. Based on a study involving weevils, it was proposed that the quorum sensing system in *S. praecaptivus* might play a key role in allowing the bacterium to establish asymptomatic infections in a broad range of insect hosts [27]. This would be anticipated to be adaptive under circumstances in which these bacteria are transmitted (between plant and/or animal hosts) by insect vectors. In addition, it was proposed that it might underlie the propensity for these bacteria to adopt, permanent, mutualistic associations with a wide range of insect hosts. In order to evaluate the latter hypothesis, we sought to determine if quorum sensing plays an important role in modulating the virulence of *S. praecaptivus* towards tsetse flies, which are representatives of an order of insects (Diptera) that is distantly related to weevils (Coleoptera). To determine the importance of quorum sensing for the successful establishment and persistence of *S. praecaptivus* within tsetse, several mutant strains with disruptions in key quorum sensing genes including *ypeI, ypeR*, and *yenR* were introduced into tsetse flies and their survival was monitored over time. A quadruple mutant lacking *ypeI*, along with the insecticidal binary toxin genes *pirA* and *pirB* [33], and a regulatory gene, *regC*, that seems to enhance *pirAB* expression (based on transcriptomic analyses) [27], was also introduced into tsetse to further assess the relationship between quorum sensing, the production of virulence factors, and tsetse fly survival.

Microinjection of a mutant lacking the gene encoding the homoserine lactone synthase, *ΔypeI*, caused a significant increase in fly mortality in comparison to flies microinjected with WT *S. praecaptivus* (Log-rank test; $p = 0.0009$, Fig. 2). Further, flies that were challenged with mutants lacking response regulators, *ΔypeR* or *ΔyenR*, exhibited even greater mortality in a shorter span of time in comparison to flies injected with WT *S. praecaptivus* (Log-rank test; $p < 0.0001$). However, the most deleterious effect on fly survival arose from a microinjection of a *S. praecaptivus* double mutant in the response regulators *ΔyenR* and *ΔypeR* (Log-rank test; $p < 0.0001$, Fig. 2) with exacerbated mortality exemplified by no tsetse fly surviving past 10 d post introduction. In contrast, flies injected with the *ΔypeIΔpirABΔregC* quadruple mutant showed enhanced survival in comparison to the isogenic *ΔypeI* mutant alone (Log-rank test; $p = 0.05$). This "rescue" effect, occurring when virulence genes that are regulated by quorum sensing are deleted, closely resembles the results obtained in the previous study on weevils [27], suggesting that the quorum sensing that *S. praecaptivus* uses to control insect infection and the toxins that engender killing may have wide applicability in insect hosts and, intriguingly, may also provide novel broad spectrum candidates for pest control. In addition, "rescue" from killing was also observed when a *ΔypeI* mutant was co-injected with an equivalent amount of WT *S. praecaptivus* (See Appendix A, Fig. 7; Log-rank test; $p < 0.0001$). This can be explained as a consequence of the WT strain producing sufficient OHHL to induce quorum in the *ΔypeI* mutant strain, preventing it from producing the virulence factors that normally lead to host killing. To further disentangle the mechanisms operating to kill tsetse when quorum sensing genes are perturbed, *S. praecaptivus* abundance was determined in flies inoculated with either WT, *ΔypeI* or the *ΔypeIΔpirABΔyregC* quadruple mutant (See Appendix A, Fig. 8). Flies injected with the *ΔypeI* mutant had a significantly greater abundance of *S. praecaptivus* relative to those injected with the WT strain (ANOVA, Tukey's multiple comparisons test; $p = 0.002$), consistent with increased pathogenesis.

In support, flies injected with *ΔypeIΔpirABΔyregC* mutants had *S. praecaptivus* densities that were in between those observed for WT and *ΔypeI* mutant strains. However, it is also possible that inappropriate regulation of QS virulence genes is causing pathogenesis, independent of (or in conjunction with) the increase in growth.

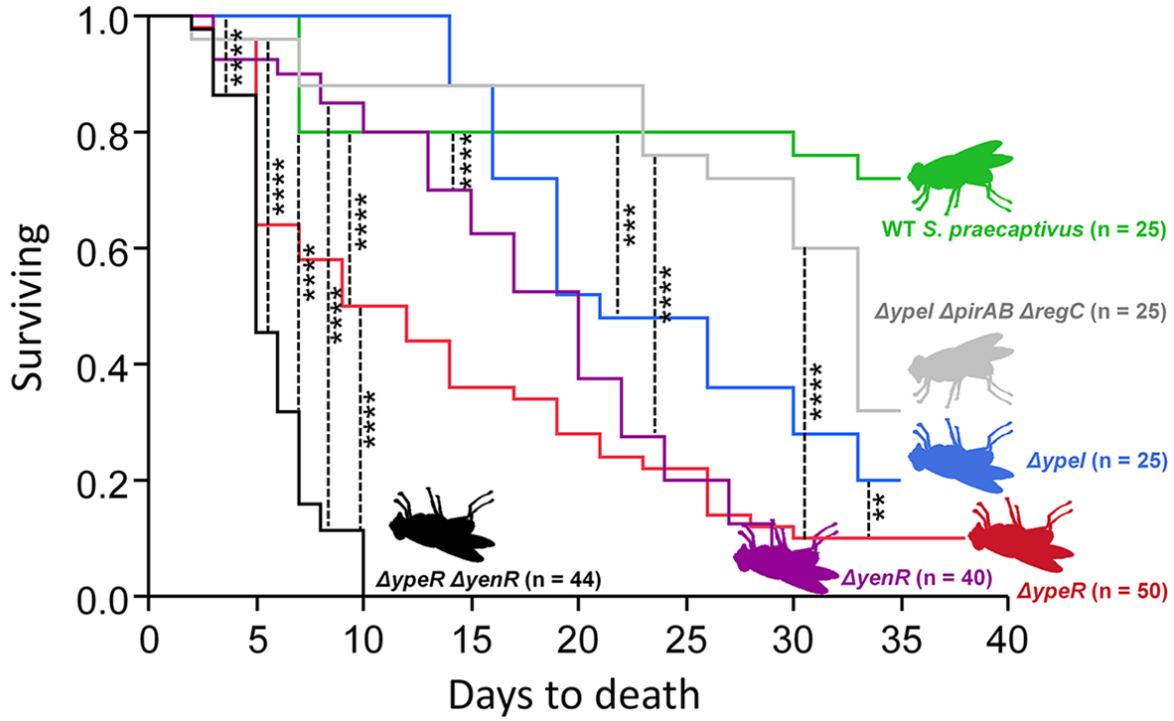

Figure 2. The disruption of *S. praecaptivus* quorum sensing (QS) genes decreases tsetse survival upon infection.
Kaplan-Meier curves comparing survival of tsetse lines injected with *S. praecaptivus* mutants in QS genes. Survival curve of a tsetse line injected with wildtype *S. praecaptivus* is included for reference; log-rank test *p*-values are indicated for the corresponding pairwise comparisons between curves. Only statistically significant differences in tsetse survival upon microinjection of *S. praecaptivus* lines are indicated with asterisks signifying; ** $p \leq$ 0.01, *** $p \leq$ 0.001, and **** $p \leq$ 0.0001. WT alone: WT *S. praecaptivus*; *ΔypeR*: mutant of the LuxR-like response regulator gene *ypeR*; *ΔypeI*: mutant with an inactivation of the OHHL synthase gene *ypeI*; *ΔyenR*: mutant of the response regulator gene *yenR*; *ΔypeR ΔyenR*: mutant in both the response regulator genes *ypeR* and *yenR*; *ΔypeI ΔpirAB ΔregC*: quadruple disruption of OHHL synthase gene *ypeI*, homologs of the *pirA* and *pirB* genes coding for a binary insecticidal toxin, and the homolog of a transcriptional regulator of *pirAB*, *regC*. n = number of flies injected per treatment. WT *S. praecaptivus* used was CD14 (Table 2 in Appendix A).
https://doi.org/10.1371/journal.pgen.1008992.g002

The impact of *S. glossinidius* towards *S. praecaptivus* prevalence and density within tsetse
We were also interested in determining the impact of autochthonous *S. glossinidius* symbionts towards *S. praecaptivus* establishment within tsetse flies, particularly because *S. glossinidius* is known to have retained the capability to synthesize OHHL [34]. The specific removal of *S. glossinidius* (and not the obligate mutualist *Wigglesworthia glossinidia*) was achieved by maintaining tsetse flies on blood supplemented with streptozotocin (2-deoxy-2-(3-methyl-3nitrosoureido)-D-glucopyranoside) [35]. Streptozotocin is a bactericidal analogue of *N*-acetyl-D-

glucosamine, the principle carbon source used for *S. glossinidius* growth within tsetse [36]. Progeny deposited by streptozotocin flies were verified to be *S. glossinidius*-free and used for the microinjection of *S. praecaptivus* (See Appendix A, Fig. 9). The absence of endogenous *S. glossinidius* populations did not affect the infection prevalence of tsetse with *S. praecaptivus* when introduced either *per os* or via microinjection. Similar infection rates were observed when comparing the prevalence of *S. praecaptivus* colonization following microinjection into *G. morsitans*^{WT} (harboring *S. glossinidius*) or into individuals lacking *S. glossinidius* (*G. morsitans*^{STZ}) (100% versus 89%, respectively; Table 1). The *per os* introduction of *S. praecaptivus* into these two populations of flies also yielded similar establishment rates (80% versus 75% with a 10^3 CFU/fly introduction and 87% versus 100% with a 10^5 CFU/fly introduction, respectively; Table 1). These results indicate that endogenous *S. glossinidius* neither facilitates nor impairs *S. praecaptivus* colonization of tsetse. Future studies should further explore this question of facilitation or inhibition by the endogenous *S. glossinidius* through testing a range of infectious doses, particularly at lower abundances (i.e., ID_{50}) which may enable better identification of subtle differences in *S. praecaptivus* colonization.

We then reasoned that even if *S. glossinidius* did not impact the establishment of infection, the density and distribution of *S. praecaptivus* might vary in the presence/absence of *S. glossinidius*. An increase in *S. praecaptivus* density within flies that harbored their endogenous *S. glossinidius* would suggest synergism between these two bacteria, while a decrease may point to competitive inhibition. To determine this, we compared the CFUs of *S. praecaptivus* in the midgut and carcass of the flies (defined as the whole fly minus the gastrointestinal tract from the anterior midgut to the Malpighian tubules), comparing wild type (+Sg) tsetse with streptozotocin-treated individuals lacking *S. glossinidius* (-Sg) following microinjection. Only -Sg flies harbored significantly higher *S. praecaptivus* densities in their guts relative to carcasses (Fig. 3A, Mann-Whitney U test, p = 0.0022). Surprisingly, there was a significantly greater density of *S. praecaptivus* detected in the carcasses of +Sg flies relative to -Sg flies (Fig. 3A, Mann-Whitney U test, p = 0.01). Tsetse midguts harboring *S. glossinidius* also contained higher *S. praecaeptivus* densities relative to –Sg midguts, but lacked statistical significance (Fig. 3A, Mann-Whitney U test, p= 0.20). This is difficult to explain on the basis that these two bacteria might be expected to compete for resources. The observed synergistic interaction certainly merits further exploration.

Since *S. glossinidius* and *S. praecaptivus* use the same signaling molecule (OHHL) to control their quorum sensing systems [27, 34], we examined the ability of endogenous *S. glossinidius* to complement the *S. praecaptivus ΔypeI* mutant. We hypothesized that if the *S. praecaptivus* homoserine lactone synthesis (*ΔypeI*) mutant could be complemented by endogenous *S. glossinidius* OHHL then flies that harbor *S. glossinidius* might be rescued from killing mediated by the *S. praecaptivus ΔypeI* mutant, in the same way that killing is suppressed when the *ΔypeI* mutant is co-injected with WT (See Appendix A, Fig. 7). Notably, following infection with the *S. praecaptivus ΔypeI* strain, female flies harboring *S. glossinidius* had a significantly longer lifespan then individuals lacking *S. glossinidius*, (Fig. 3B, 46.5 $\pm$ 2.1 d versus 24.8 $\pm$ 1.2 d, respectively, Mann-Whitney U test, $p < 0.0001$). However, this effect was not observed in male flies, with members of both the -Sg and +Sg groups showing a similar mean longevity (Fig. 3B, 19.9 $\pm$ 3.4 d versus 22.3 $\pm$ 6.0 d for +Sg males and -Sg males; respectively, Mann-Whitney U test, p =0.6825). This difference may be explained by the fact that female tsetse harbor significantly higher densities of *S. glossinidius* [37], perhaps yielding sufficient (*S. glossinidius*-derived) OHHL to suppress killing

by the *ΔypeI* mutant strain. Further, it is possible that the difference in survival observed between the sexes may be a function of the different lifespans of males and females [38], where the significantly shorter life span of males limits our ability to detect a rescue effect.

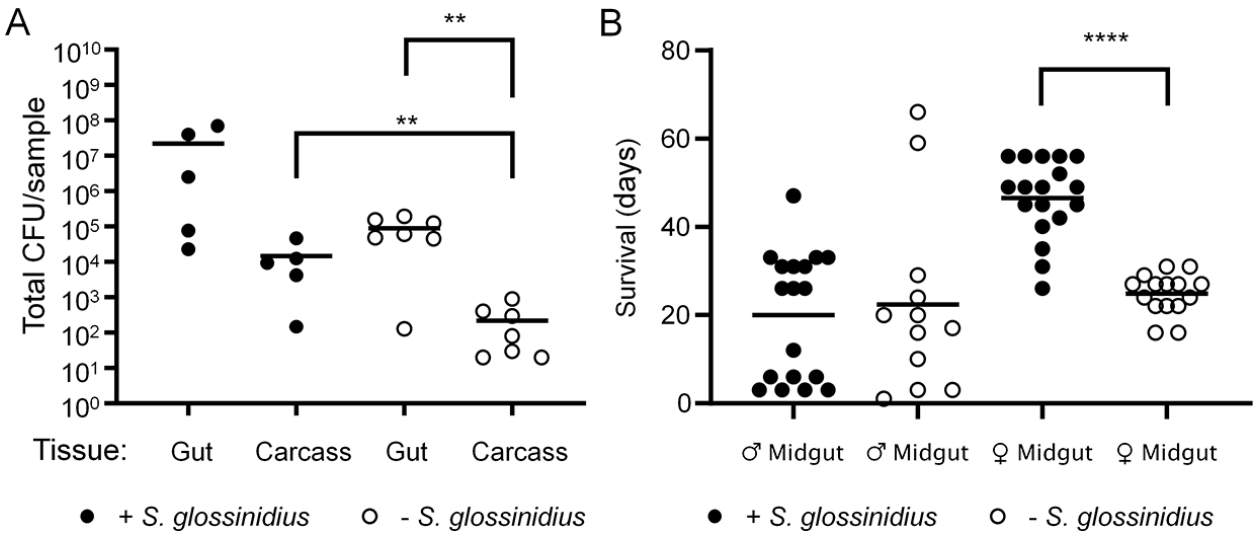

Figure 3. The effects of *S. glossinidius* on *S. praecaptivus* density and distribution within the tsetse fly and towards the complementation of the *S. praecaptivus ypeI* mutant.
A. Comparison of *S. praecaptivus* abundance and distribution in flies seven days post introduction, with or without autochthonous *S. glossinidius*. All groups were injected with WT *S. praecaptivus*. The density of *S. praecaptivus* is significantly higher in the gut, relative to the carcass, when *S. glossinidius* is lacking within tsetse. The density of *S. praecaptivus* was significantly higher in tsetse carcass that harbored native *S. glossinidius* in comparison to flies that lacked this endosymbiont. Horizontal lines represent mean CFU per treatment, while dots indicate the bacterial CFU within individual flies. B. Endogenous *S. glossinidius* complements *S. praecaptivus ΔypeI* within females. Fly groups were injected with *S. praecaptivus ΔypeI* mutant and lifespan assessed per treatment group. Females showed a significant increase in lifespan when *S. glossinidius* is present upon the introduction of *S. praecaptivus ΔypeI* mutant. Horizontal lines represent mean survival (days) per treatment, while dots indicate the survival of individual flies, ** $p \leq 0.01$ and **** $p \leq 0.0001$. WT *S. praecaptivus* used was CD623 (Table 2 in Appendix A).
https://doi.org/10.1371/journal.pgen.1008992.g003

S. praecaptivus infections impact tsetse fecundity

While WT *S. praecaptivus* infections persist long-term with no detectable impact on tsetse lifespan, we were interested in examining the effect of these infections on tsetse reproductive fitness. We first examined the impact of both parents harboring *S. praecaptivus* infections towards tsetse reproductive fitness since transmission of *S. glossinidius* can occur from either parent [39-42]. To accomplish this, both female and male tsetse were microinjected with 1 x 10^5 WT *S. praecaptivus* and subsequently mated. The number of larvae deposited per mated female was determined weekly and compared to the fecundity of control flies (Fig. 4A). Both *S. praecaptivus* infected and uninfected females began depositing pupae approximately one week following mating, indicating no lag in time to offspring deposition between the treatment groups. Further, comparison of the regression lines through the weekly fecundity of *S. praecaptivus* inoculated flies (Y= -0.001838*X + 0.1835) with that of controls (Y= 0.0009411*X + 0.4968), demonstrated no significant difference between slopes (Multiple linear regression, p= 0.9194), reflecting a constant reproductive rate in both groups. However, the number of larvae deposited per female was

consistently higher in controls as supported by a significantly higher Y intercept for this group (Multiple linear regression, p=0.0257). Despite a lower reproductive output, pupae deposited by the *S. praecaptivus*-infected parental line were similar in weight (0.02571 ± 0.0014g) to those deposited by the control group (0.02471 ± 0.0006g; Mann-Whitney test, p = 0.7466; Fig. 4B). These results indicate that while tsetse flies inoculated with *S. praecaptivus* infection do reproduce successfully and produce viable and phenotypically normal offspring, their reproductive output is significantly lower, perhaps as a consequence of increased use of host (nutritional) resources by *S. praecaptivus*. Whether these *S. praecaptivus* infection costs are transgenerational merits further studies.

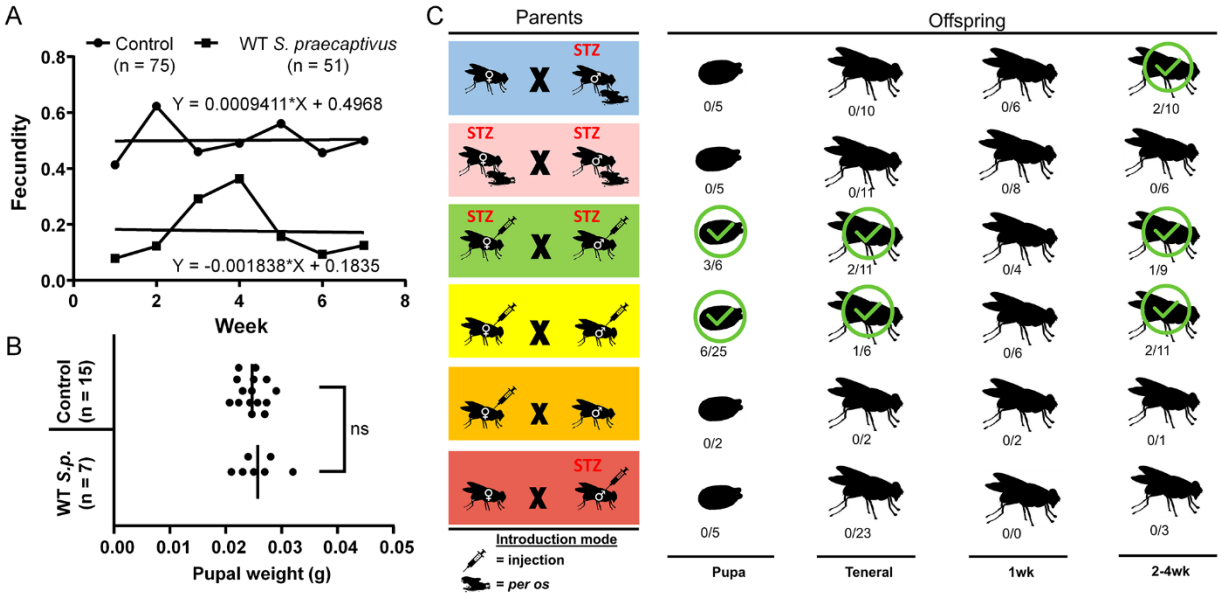

Figure 4. *S. praecaptivus* is vertically transmitted in tsetse but infection incurs a reproductive output cost.
A. WT *S. praecaptivus* infected flies are fertile. Multiple linear regression analysis shows that fecundity of both control (n = 75) and infected (n = 51) flies commence at equivalent times remaining consistent through time, however; the reproductive output of flies infected with *S. praecaptivus* is lower. B. There was no significant difference in the weight of pupae deposited by *S. praecaptivus* infected flies upon comparison to pupae deposited by age matched control flies. Vertical bars indicate the mean pupal weight (grams) within a treatment group, each dot represents an individual pupa. C. Infection status at different life stages of progeny arising from different crosses of tsetse individuals infected with *S. praecaptivus*. The presence of viable *S. praecaptivus* within progeny was verified via plating individual fly homogenates in selective media with identity confirmed through *tam* specific PCR. For each cross, the number of progeny positive for *S. praecaptivus*/ total number of progenies sampled at that time point are provided. STZ, streptozotocin-treated. WT *S. praecaptivus* used was CD623 (Table 2 in Appendix A).
https://doi.org/10.1371/journal.pgen.1008992.g004

S. praecaptivus is vertically transmitted in tsetse

We were interested in examining whether following establishment within tsetse, *S. praecaptivus* would undergo vertical transmission, a feature that is necessary for evolutionary persistence and/or development of mutualistic associations. To test the potential for vertical transmission within tsetse, WT *S. praecaptivus* was either injected or fed to females and/or males prior to mating. The resulting offspring were then tested for *S. praecaptivus* colonization at different points in pupal

and adult development. The vertical transmission of *S. praecaptivus* was confirmed in three distinct tsetse lines at various times during development (Fig. 4C); 1) males^{STZ} fed *S. praecaptivus* x females^{WT} (6.5% of offspring), 2) males^{STZ} injected with *S. praecaptivus* x females^{STZ} injected with *S. praecaptivus* (20% of offspring) and 3) males^{WT} injected with *S. praecaptivus* x females^{WT} injected with *S. praecaptivus* (19% of offspring). Notably, each of the successful combinations includes infected males and one of those includes uninfected females, indicating that paternal transmission took place. Although no transmission was observed under any conditions in which the male in the pairing was uninfected, it is notable that the highest frequency of infected progeny arose when both parents harbored *S. praecaptivus*. We found colonization of *S. praecaptivus* within testes, ovaries and spermathecae (i.e., female receptacle used for sperm deposit) within 100% of microinjected virgin flies. In additional support for the invasion of gonads, a mCherry-expressing *S. praecaptivus* localized to tsetse ovaries, spermathecae, and testes of microinjected flies. Comparable densities were found in these reproductive tissues suggesting that multiple routes of parental inheritance may enable *S. praecaptivus* transmission within tsetse (Fig. 5). The vertical transmission of *S. praecaptivus* in tsetse is remarkable in the sense that the partners are naïve and have not been subject to an evolutionary opportunity to co-evolve. Moreover, this provides further evidence of the capacity of the *Sodalis*-allied symbionts to colonize a wide range of insects and develop mutualistic associations.

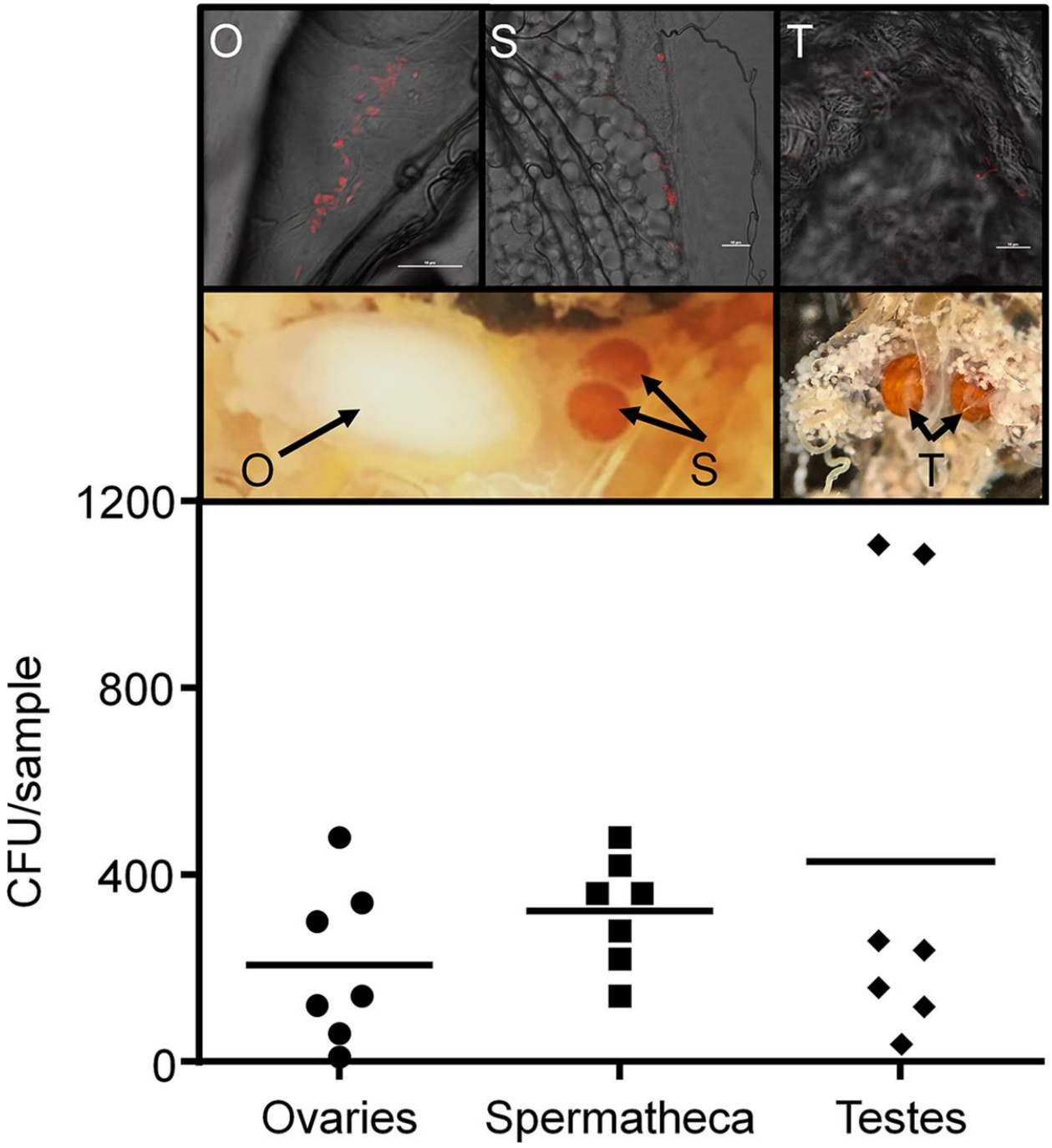

Figure 5. Colonization of tsetse reproductive tissue by *S. praecaptivus*.
Tsetse individuals (unmated) received intrathoracic injections with 1 x 10^5 WT *S. praecaptivus* (CD623, Table 2 in Appendix A) and 7 d post introduction reproductive tissues were dissected. Infections occurred in all sample types with no significant differences in density (ANOVA, $p > 0.05$). Horizontal lines represent mean CFU per treatment, while each dot represents the CFU within an individual. Insets above show mCherry-expressing *S. praecaptivus* invading the reproductive tissue of both sexes. O; ovary, S; spermathecae, T; testis. Scale bar is 10 μm.
https://doi.org/10.1371/journal.pgen.1008992.g005

Genes under quorum sensing control in both *S. glossinidius* and *S. praecaptivus*

Only truncated versions of *cpmA* (orthologs encode carbapenam-3-carboxylate synthase) and *cpmJ* (orthologs are involved in the resistance to the antibiotic carbapenem) are found in *Sodalis spp*. (See Appendix A, Table 3 in Appendix A and Fig. 6A). This gene retention pattern suggests a distinct functional role from antibiotic synthesis, which could be either ancestral to, or derived from, a complete carbapenem gene cluster that is capable of producing antibiotic [34]. Notably, *cpmA* and *cpmJ* are both under quorum sensing regulation, and unlike the virulence genes that are repressed under quorum, the expression of these genes is enhanced within both *S. glossinidius* and *S. praecaptivus* [27, 34], which may indicate that it has an adaptive function in insect symbiosis.

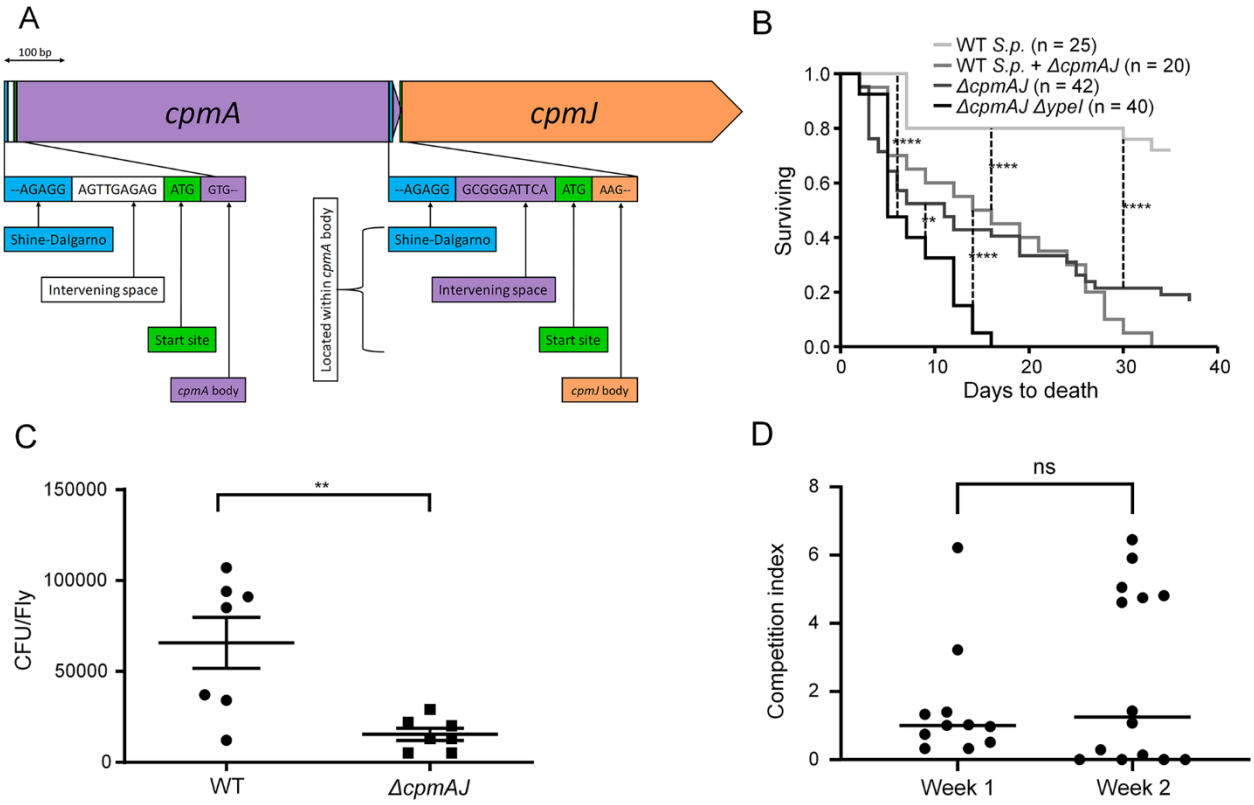

Figure 6. Homologs within carbapenem biosynthesis operon mediate attenuation of virulence in *S. praecaptivus*.

A. Structure of *S. praecaptivus* genes *cpmA* and *cpmJ* (Sant_3163 and Sant_3164). B. In *S. praecaptivus*, disruption of QS regulated genes in the carbapenem biosynthesis pathway (*ΔcpmAJ*) increases virulence leading to lower tsetse survival: ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$. WT *S. praecaptivus* used for comparison was CD14 (Table 2 in Appendix A). C. The density of *S. praecaptivus* within flies following 20 d post introduction was assessed through CFU counts after surface sterilization, homogenization, serial dilution and plating of whole flies. Horizontal lines represent mean bacterial count, while dots represent mean bacterial count per individual. Mean bacterial density was compared (Student's *t*-test, **, $p = 0.0022$). Bars represent 1 SEM. D. The *S. praecaptivus* mutant strain *ΔcpmAJ* does not outcompete the WT *S. praecaptivus* CD623. Competition index (CI) = (*ΔcpmAJ* output/WT output)/(*ΔcpmAJ* input/WT input). Horizontal lines represent mean CI between *S. praecaptivus* *ΔcpmAJ* and WT, while dots indicate the mean CI within individual flies.

https://doi.org/10.1371/journal.pgen.1008992.g006

In order to assess the importance of *cpmA* and *cpmJ* in the context of a nascent symbiosis, both genes were inactivated in *S. praecaptivus* and mutants were injected into tsetse. The survival of the *S. praecaptivus ΔcpmAJ* inoculated flies was significantly reduced in comparison with that of flies injected with the wildtype strain (Log-rank test, $p < 0.0001$; Fig. 6B). Moreover, when a *ΔcpmAJ ΔypeI* triple mutant was introduced into tsetse, the survival of flies decreased even further (*ΔcpmAJ* vs. *ΔcpmAJ ΔypeI;* Log-rank test, p = 0.0024; Fig. 6B). When *S. praecaptivus* abundance was determined in flies inoculated with either WT or *ΔcpmAJ* mutants (Fig. 6C), a significantly higher bacterial density was found in flies harboring WT (Student's *t*-test, $p = 0.0044$) indicating that host killing was occurring in the presence of only a small number of bacterial cells.

We were curious whether a WT strain of *S. praecaptivus* could reduce the virulence of the *ΔcpmAJ* strain upon mixed infection, in the same way that the WT strain can suppress the killing effect of the *ΔypeI* strain by providing a source of OHHL. However, injection of a 1:1 mixture of WT *S. praecaptivus* and the *ΔcpmAJ* mutant strain did not restore tsetse survival relative to flies injected with the *ΔcpmAJ* mutant alone ((WT + *ΔcpmAJ*) vs. *ΔcpmAJ*; Log-rank test, p = 0.5652; Fig. 6B). However, tsetse inoculated with the mixture of WT and *ΔcpmAJ* mutant *S. praecaptivus* maintained equivalent levels of both of these strains at one and two-week timepoints following injection (Fig. 6D). Finally, we also tested the effect of the *ΔcpmAJ* mutant on grain weevil survival, to determine if it had a similar effect in those insects. Similar to what we observed in tsetse flies, a mutant in quorum sensing regulation (*ΔypeR*) was even more virulent when coupled with the *ΔcpmAJ* mutation (See Appendix A, Fig. 10). This suggests that the *S. praecaptivus* CpmA and CpmJ proteins have a general protective effect in insects following infection. However, since little is known about the function of *cpmAJ* in *Sodalis* spp., further investigations are needed to determine how its activity limits virulence.

Discussion

Symbiosis follows diverse evolutionary paths

Upon entering a host, bacteria encounter specific physical and biological challenges that mandate the expression of specialized genes that facilitate bacterial survival. Establishment within a host requires sequestration of host resources for nutrition, resistance towards host immunity and (often) the secretion of virulence factors that facilitate entry of bacteria into host cells and tissues. While parasitic and mutualistic symbionts both need to establish access, overcome host defenses, and obtain nutrition, mutualists have the added burden of limiting their virulence, in order to minimize negative impact on host fitness. In addition, mutualists need to evolve mechanisms to achieve vertical transmission, to ensure their maintenance in host offspring, providing natural selection with an opportunity to optimize traits that favor the persistence of the symbiotic partnership in nature [43].

S. praecaptivus exhibits wide insect host adaptability

The current work, along with previous work, shows that *S. praecaptivus* is capable of establishing persistent infections in weevils and tsetse flies. The Coleoptera and Diptera orders belong to two major lineages of Holometabola, the Neuropteroidea and the Mecopterida, respectively. The divergence of these groups is estimated to have taken place in the Permian period, during the Paleozoic Era, around 300 million years ago [44]. Additionally, weevils and tsetse flies have widely different feeding habits, the former feeding on a variety of grains [45], while the latter consumes

blood [46]. With such stark differences in host physiology and dietary ecology, it might be anticipated that a bacterium attempting to infect these hosts would face acutely distinct physiological environments that limit its ability to infect these diverse hosts. However, *S. praecaptivus* finds suitable nutrients for survival and acclimates to the microenvironmental physical constraints imposed by these divergent insect hosts. This highlights the potential of this bacterium to establish infection in a wide range of insect hosts, concordant with the observation that *Sodalis*-allied symbionts have been identified in a wide range of blood-feeding and plant-feeding insect taxa[25-27].

Our work shows that wild-type *S. praecaptivus* establishes a stable association in tsetse flies, as it does in grain weevils [27], characterized by a longstanding infection that has no detectable impact on insect longevity. However, our work also shows that *S. praecaptivus* mutants, lacking key components of the quorum-sensing machinery, have a potent insect killing phenotype that is correlated with the loss of repression of virulence gene expression that is normally mediated by their quorum sensing gene regulatory system. Interestingly, host killing is mediated, at least in part, by PirAB toxins, which have been studied in other bacteria (notably *Photorhabdus spp.*) [33], but were not known to have such broad insecticidal activity.

Quorum sensing attenuates virulence

The potential to infect a wide assortment of animal hosts is explained by the fact that the *S. praecaptivus* genome harbors many genes that are predicted to encode virulence factors and toxins believed (by virtue of homology) to target insect, plant, and mammalian hosts. In terms of insect infections, specific factors including PirAB seem to serve to facilitate host establishment but are then suppressed transcriptionally, via quorum sensing based regulation, to enable bacterial persistence without negatively impacting host fitness [27]. Following establishment and vertical transmission, degenerative evolution is expected to catalyze inactivation and loss of genes that no longer have adaptive value in these bacterial-insect associations [25, 47, 48]. These adaptations evolve in accordance with functional mandates that are specific to a given host and are therefore anticipated to result in an ecological tradeoff that should also restrict the transfer of established symbionts into novel host species.

Virulence is a context-dependent trait that is influenced by bacterial infection dynamics [49] and the quantitative and qualitative mandate for toxin production [50]. On a basic level, parasites are anticipated to exhibit increased virulence relative to mutualists because their hosts are dispensable. On a more subtle level, mutualists are anticipated to minimize their virulence (and hosts are anticipated to exhibit immunotolerance towards their mutualistic microbial partners [51]), so that energy is not wasted in unnecessary conflict. Thus, it makes sense that while genes encoding quorum sensing-regulated, insect-specific virulence factors (including *pirAB*) are maintained in a functional state in *S. praecaptivus* [27], their homologs are often found to be inactivated or lost in the genomes of the recently established mutualistic symbionts *S. glossinidius* [31] and *Ca.* S. pierantonius [47], both of which retain a functional quorum sensing system [34]. However, it is notable that both of those insect symbionts retain identifiable homologs of *cpmA* and *cpmJ*, which are among the few genes whose expression was demonstrated to be upregulated under quorum in *S. praecaptivus* [27] and *S. glossinidius* [34]. Strikingly, in this study, we show that the *cpmAJ* genes appear to have a beneficial (protective) function during *S. praecaptivus* infection of both weevils and tsetse flies but likely in distinct ways. Although the mechanism of this protective effect is not

yet understood, it is interesting to note that tsetse flies inoculated with *ΔcpmAJ* mutants displayed increased mortality in the absence of significant bacterial proliferation. Based on the observation that the homolog of CpmJ found in carbapenem-synthesizing bacteria (CarG; an intrinsic carbapenem resistance protein) may function by modifying penicillin-binding proteins in the bacterial cell wall [52], it is conceivable that CpmAJ could modulate interactions between *S. praecaptivus* and the insect immune system, increasing host fitness [53]. In contrast, the *cpmAJ* mutant does not compromise survival, and bacterial density is comparable to wildtype *S. praecaptivus* densities within weevils. However, when we examined Δ*ypeR* Δ*cpmAJ*, we saw acceleration of host killing compared to Δ*ypeR* in weevils. This result indicates CpmAJ has a host protective effect in weevils as well. At this time, we do not know enough to speculate on the different roles of CpmAJ within distinct hosts.

The ecological mandate for quorum sensing mediated virulence suppression

Our results demonstrate that the quorum-sensing mediated control of *S. praecaptivus* virulence in tsetse flies operates in a remarkably similar way to that observed in grain weevils [27]. Based on the fact that these insects are so distantly related, it is tempting to speculate that quorum sensing mediated control of virulence would work in a wide range of insect hosts, explaining (at least in part) the widespread distribution of *Sodalis*-allied symbionts among insect taxa. However, because natural selection lacks foresight, it is important to recognize that the quorum sensing based modulation of virulence in *S. praecaptivus* must play an important (adaptive) role in the ecology of this bacterium in its free-living state. To this end, it has been suggested that *S. praecaptivus* is an opportunistic pathogen of plants and animals that may develop associations with insects in order to achieve vector-based transmission [25, 27]. This is consistent with the observation that *S. praecaptivus* utilizes quorum sensing to limit its virulence towards insects based on the intuitive notion that natural selection will favor passengers that minimize the fitness costs associated with their transmission by vectors.

Insight into the origins of *Sodalis*-insect associations

Because vectors will always be negatively impacted by the metabolic cost of maintaining a bacterial passenger, a trade-off might be anticipated to evolve in which the passenger provides the vector with some benefit that offsets the cost of maintenance, thereby sowing the seeds of a mutualistic relationship. Under circumstances in which there is a net benefit to the host, natural selection would then be expected to favor the evolution of a vertical transmission strategy [54]. In the case of many insect-bacterial mutualisms, this is not difficult to rationalize on the basis that many free-living microbes are capable of synthesizing essential amino acids and vitamins that cannot be synthesized by insects and are lacking in certain diets such as blood or plant sap. As an alternative outcome to mutualism, other interactions may evolve such as parasitism driven by the enhanced fitness of the symbiont through host manipulation as observed with reproductive parasites.

Strikingly, our results show that, in addition to having a normal lifespan, tsetse flies infected with *S. praecaptivus* are capable of transmitting these bacteria vertically to their offspring (albeit at a relatively low frequency). Transmission events that were detected in our experiments always involved an infected male and could take place in the absence of an infected female. This indicates that transmission must have occurred via a strictly paternal route, as has been reported for *S. glossinidius* [42], but does not rule out the possibility that transmission can also occur via a maternal

route. Moreover, both WT and mCherry-expressing *S. praecaptivus* cells were detected in tsetse ovaries, testes, and spermathecae.

While more experimentation is needed to understand the dynamics of *S. praecaptivus* transmission in tsetse, our experiments do reveal that *S. praecaptivus* infection does incur a significant cost at the level of tsetse reproduction. Thus, in the absence of a significant compensatory benefit, flies infected with *S. praecaptivus* would be anticipated to be outcompeted by non-infected individuals. Obviously, our experiments were conducted in a laboratory setting in which there is no mandate for an adaptive benefit resulting from *S. praecaptivus* infection. However, in nature, associations are predicted to arise within the context of an opportunity for niche expansion. In addition, it stands to reason that the maintenance of a novel bacterial infection, with no adaptive benefit, would incur a fitness cost. Indeed, the acquisition and transmission of a novel bacterial partner are only expected to be beneficial if it facilitates adaptation towards environmental change, such as dietary niche expansion or resistance to a novel pathogen.

Clearly, the *Sodalis*-allied symbionts have been very successful in gaining entry into insects and evolving various symbiotic roles. However, we still understand little about how these associations originate in nature. Our results indicate that infection in tsetse flies can occur through oral introduction or hemocoel injection. Comparative genomic studies suggest that the establishment of *Sodalis*-allied symbionts has occurred independently in many insect hosts and our results certainly provide support for that notion. They also indicate that persistent infection can be maintained in the presence of existing *Sodalis*-allied symbionts, including *S. glossinidius* (in our study), where synergistic interactions between the two bacterial species may be occurring. For example, the production of OHHLs by *S. glossinidius* may enhance the number of *S. praecaptivus* in tsetse, although this remains speculative. This supports the notion that novel symbiont acquisitions could be added to augment the functions of existing symbionts, or that symbiosis can persist as a "revolving door", in which existing symbionts that have been functionally compromised (by degenerative mutations) are refreshed by free-living candidates. Considering both of these scenarios, it seems important for those candidates to be able to persist and be transmitted efficiently. To that end, our work shows that *S. praecaptivus* has the ability to undergo vertical transmission in tsetse, albeit in a limited capacity and with negative impacts on host reproductive fitness. While we propose that this capability may help to explain the propensity of *Sodalis* spp. to develop associations with a wide range of insects, as observed in nature, it is important to recognize that natural selection has no ability (foresight) to act on future outcomes. Therefore, the ability of *S. praecaptivus* to undertake a cautious and controlled infection of insect hosts must have an adaptive benefit in the lifecycle of this bacterium prior to the evolution of stable associations.

Application of *S. praecaptivus* to symbiosis research and paratransgenesis

Because of the ease of culture and genetic manipulation in the laboratory and its ability to infect a wide-range of insect hosts, *S. praecaptivus* represents an excellent candidate to study mechanistic aspects of symbiosis. The discovery that it is vertically transmitted in tsetse opens the door to discover and study the determinants of that mode of transmission through genetic and cell biological approaches. It also provides an opportunity to engineer *S. praecaptivus* strains that have increased efficiency of transmission and that phenocopy existing symbionts, such as those found in tsetse flies. This is particularly interesting in the context of paratransgenesis in which symbionts

could be used to express genes that prevent insect vectors (such as tsetse, obligate vectors of African trypanosomes) from transmitting diseases of medical and agricultural importance. In the case of tsetse, such technology is envisioned to be used to augment the sterile insect technique (SIT), in which irradiated male flies are released in the wild to mate with wild females, which then produce no offspring. One of the challenges with the implementation of SIT is that the release of the males can lead to an increase in disease transmission because the sterile males are still competent to vector trypanosomes. Thus, it has been proposed that paratransgenic male tsetse flies, harboring symbionts that express anti-trypanosomal gene products, could be used to ameliorate this negative consequence [55]. Another challenge in this context is that the symbionts themselves need to be able to withstand insect irradiation [56]. To this end, while *S. glossinidius* lacks a number of key DNA repair enzymes [27], *S. praecaptivus* naturally retains a full complement of this machinery, suggesting that it would be more robust in this application. Further, it should be noted that efficient vertical transmission is not a prerequisite for the implementation of this approach. Indeed, recombinant *S. praecaptivus* could be introduced to tsetse by pupal or adult microinjection prior to release.

Conclusion

Our findings shed light on the molecular interplay taking place in the incipient stages of symbiosis, highlighting the versatility of quorum sensing towards facilitating adaptability and persistence of *S. praecaptivus* within dissimilar insect hosts. The novelty of our results is first rooted in providing support for *S. praecaptivus* being emblematic of a proto-symbiont for the *Sodalis* clade, particularly given its capability towards establishing in a wide host range. Secondly, we present evidence that quorum sensing is essential for transitioning from an environmental to a symbiotic state irrespective of the target host. Third, we demonstrate that the regulation of virulence through quorum sensing is fastened to an unprecedented ability to undergo vertical transmission quickly within a single host generation. Importantly, given the amenability towards genetic manipulation and its vertical transmission within tsetse, further optimization of *S. praecaptivus* towards paratransgenesis efforts could be used to bolster vector control strategies.

Materials and methods

Tsetse lines. The tsetse species, *Glossina morsitans*, is maintained in WVU's insectary at 24°C with 50-55% relative humidity on a 12 h light:12 h dark photoperiod schedule. All flies receive defibrinated bovine blood (Hemostat, CA) every 48 h through an artificial membrane feeding system [57].

***Sodalis praecaptivus* cultures and genetic manipulation.** The *S. praecaptivus*^{WT} and mutant strains (See Appendix A, Table 2) were grown overnight with shaking at 200 rpm at 30ºC in LB media (lacking NaCl) with antibiotics as necessary [27]. When needed, antibiotics were added to media at the following concentrations: spectinomycin 40 µg/mL; gentamicin 5 µg/mL, kanamycin 25 µg/mL, tetracycline 5 µg/mL. Disruptions/deletions in *S. praecaptivus* genes were generated with the Lambda red recombineering system as reported previously [27].

***Sodalis praecaptivus* mCherry mutant strain.** The strain expressing mCherry was constructed using the Lambda red recombineering methods reported previously [27] using a construct comprising

a Zeocin resistance cassette linked to an mCherry allele that was codon-optimized for expression in gamma-Proteobacteria [58, 59].

Introduction of *S. praecaptivus* into tsetse flies. The introduction of *S. praecaptivus* into tsetse flies was performed either through *per os* (i.e., oral introduction) into tenerals (i.e., newly emerged adults prior to acquiring a blood meal) or by microinjection of individuals following one blood meal to enable hardening of the cuticle prior to puncture. The introduction of *S. praecaptivus per os* involved feeding flies a heat-inactivated (56ºC for 30 min to destroy complement) blood meal which resulted in inoculations of 1.0×10^3 or 1.0×10^5 bacteria per fly. Flies that did not take a blood meal, determined through visual observation of tsetse abdomens, were excluded from analyses. The viability of the bacteria within heat-inactivated blood was confirmed by incubation for one week under the same conditions used to rear tsetse and then inoculating on L-plates (LB + 40 μg/mL X-Gal + 7 μg/mL IPTG, No NaCl) to confirm the presence of *S. praecaptivus* based on colony morphology and expression of ß-galactosidase. For microinjection, flies were first immobilized on ice and injected into the midthoracic region with needles immersed into overnight bacterial cultures. Microinjection resulted in approximately $1x10^5$ *S. praecaptivus* cells being introduced into each fly, as verified by plating needle contents. Following *S. praecaptivus* introduction, tsetse were returned to colony rearing conditions and fed defibrinated bovine blood with mortality recorded every 48 hours.

The establishment and density (CFU) of *S. praecaptivus* infections within tsetse was determined by surface sterilizing flies in 10% bleach for 5 min, homogenizing individuals in sterile deionized water, and plating a dilution series of the homogenates on L-plates. Colonies were confirmed to be *S. praecaptivus* through PCR amplification of the trans-aconitate 2-methyltransferase (*tam;* GenBank: AHF76984.1) gene using primers *tam*127 (5'-GCT ATT GGT CGA GCG TTT TAC C-3') and *tam*128 (5'-CGG CAT CAC ATG GTA ATA GC-3') with a 58°C annealing temperature.

Tsetse reproductive fitness and *S. praecaptivus* infection. Following one blood meal, female and male tsetse received a thoracic microinjection of 1×10^5 WT *S. praecaptivus*. Following 7 d post *S. praecaptivus* injection challenged flies were mated with mortality and pupal deposition recorded every 48 hours and compared to age-matched mated controls. Tsetse reproductive fitness through time was determined by dividing the number of pupae deposited weekly by the number of mated females alive during each observation period.

***S. praecaptivus* vertical transmission.** Pupae obtained from mating crosses of *S. praecaptivus* challenged parents (as described above) were collected, weighed and reared individually. Pupae were randomly sampled through development to test for a viable *S. praecaptivus* infection by surface sterilization, homogenization, and plating on selective media as described above. Colonies were verified as *S. praecaptivus* through PCR amplification of the *tam* gene.

Please refer to *Appendix A*, Materials and Methods for more information regarding generating *S. glossinidius*-free tsetse lines and the *S. praecaptivus* mCherry mutant strain. Information on fluorescence microscopy, competitive index, and statistical analyses may be found in the *Appendix A.*

Data availability. All data are contained in the main text and *Appendix A*. *S. praecaptivus* cultures can be obtained through ATCC (BAA-2554). Specific *S. praecaptivus* mutants are available upon request.

Acknowledgements

We thank Adam Pollio, Dylan Richmond and Ying Zhang for their technical and insectary assistance and Yinghua Su and Li Szhen Teh who engineered the *S. praecaptivus* mCherry strain. We thank Prof. Philippe Cluzel (Harvard Univ.) for providing of the mCherry-zeocin resistance cassette and Dr. Amanda Ammer (WVU) with assistance in microscopy. Imaging experiments and analyses were performed at the WVU Microscope Imaging Facility that has been supported by the WVU Cancer Institute, the WVU HSC Office of Research and Graduate Education, and NIH grants P20RR016440, P30GM103488, P20GM103434, and U54GM104942.

Appendix A

Additional materials and methods

Tsetse flies lacking *S. glossinidius*. Tsetse flies that were cleared of endogenous *Sodalis glossinidius* were generated by collecting the pupae from a parental line maintained on streptozotocin (20 µg/mL) supplemented blood meals [1]. Random tsetse individuals (n=10) from the streptozotocin treated parental line are checked at each generation to verify the absence of *S. glossinidius* through PCR of whole fly DNA using oligonucleotides that amplify an internal transcribed spacer (ITS) region as described previously in [2]. *Sodalis*-free tsetse flies are maintained on separate feeding membranes from wild type flies.

***Sodalis praecaptivus* mCherry mutant strain.** The strain expressing mCherry was constructed using the Lambda red recombineering methods reported previously [3] using a construct comprising a Zeocin resistance cassette linked to an mCherry allele that was codon-optimized for expression in gamma-Proteobacteria [4, 5]. This construct was modified for recombination by adding flanking sequences that target the lacZ gene in *S. praecaptivus*, based on the notion that lactose utilization is dispensable for insect association. The mCherry-zeocin cassette was first amplified by PCR in a reaction composed of 10µl 5X PCR buffer (Thermo Fisher Scientific), 4µl 25mM dNTPs, 3µl 25mM $MgCl_2$, 1.25µl of 20µM forward primer (5'-CTTCTTAAACATAAAGTGTCTC-3'), 1.25µl of 20µM reverse primer (5'-GACAGTCATTCATCTTTCTGC-3'), 0.5µl of Phusion High Fidelity DNA Polymerase (Thermo Fisher Scientific) and 1µl of purified template DNA. The thermocycling was performed using a 2 min denaturation at 98°C followed by 29 cycles of 98°C for 30 s, 56°C for 30 s, 72°C for 1 min, and a final extension at 72°C for 2 mins.
To generate the recombination construct, 212 bp of the 5' end (5'-AAGTCACACGCTCACACCAG-3'/5'-GTTTATAAGGAGACACTTTATGTTTAAGAAGACGTGGTTGCACGTAAATGA-3') and 278bp of the 3' end (5'-CTTTTGGAGGGGCAGAAAGATGAATGACTGTCTTGACCGAGACAGCTCATTG-3'/5'-TCAGCATCGCAGTCTTCATC-3') of the *S. praecaptivus lacZ* gene were amplified with flanking tails using the following PCR reactants: 12.5µl of 2X Phusion PCR buffer, 6.5µl of nuclease free water, 2.5µl of 2.5µM forward primer, 2.5µl of 2.5µM reverse primer and 1µl of *S.*

praecaptivus WT DNA. The PCR was performed using an initial denaturation of 98°C 30 s followed by 35 cycles of 98°C for 10 s, 58°C for 30 s and 72°C for 2 min. The resulting PCR products were then purified using Agencourt AMPure XP beads, according to the product instructions. Four microliters of the purified products were then combined with 4ul of the mCherry-Zeo^R amplicon and 12µl of 2X Taq Polymerase MasterMix (Thermo Fisher Scientific) and subject to initial denaturation at 95°C for 30 s followed by 10 cycles of 94°C for 15 s, 45 °C for 30 s and 72 °C for 1 min. The final desired recombination construct was amplified with 1µl each of forward (5'-AAGTCACACGCTCACACCAG-3') and reverse primers (5'-TCAGCATCGCAGTCTTCATC-3'), 25 µl of 2X Taq Polymerase MasterMix (Thermo Fisher Scientific), and 23 µl of the PCR product from the 10-cycle reaction using an initial denaturation at 95°C for 30 s, followed by 35 cycles of 94°C for 15 s, 58°C for 30 s and 72°C for 1.5 min. The PCR product was then purified using Agencourt AMPure XP beads.

For Lambda-Red recombineering, 2.5 ml of an overnight *S. praecaptivus* strain 101 culture which maintains a plasmid with inducible Lambda-Red recombination functions was grown in 25 ml 2YT 5.8 media (20 mg/mL Tryptone, 8 mg/mL Yeast Extract, 10 mg/mL NaCl, pH 5.8) with 25 µg/ml chlorophenicol for 8 h at 30 ºC. The expression of the Lambda-Red functions was then induced by adding 500 µl of 20% arabinose to the culture and allowing it to grow for another 30 min. The bacteria were then centrifuged (8,000 X g, 5 min) and washed in cold sterilized DI water twice to make them electrocompetent. The 50 µl of prepared disruption fragment was then transferred into the electrocompetent cells by electroporation at 1600V/s using an Eppendorf electroporator model 2510. The cells were permitted to recover for 8-10 h recovery on an L plate at 30 ºC and then replica plated onto an L plate with 15 µg/ml zeocin and IPTG/X-gal to select for lac disruption (i.e., lac$^-$, zeocin^R). Recombinant colonies demonstrated a pink appearance as a consequence of *lacZ* inactivation and mCherry expression. They were then validated by PCR and sequencing and subjected to fluorescence microscopy to ensure mCherry expression.

Fluorescence microscopy. Gonads and spermathecae from virgin flies 7 d following microinjection of *S. praecaptivus* mCherry mutant strain were dissected in Phosphate Buffered Saline and placed on microscope slides. Images were obtained using an inverted Nikon A1R confocal with a 60X/1.4 plan Apo oil objective. The *S. praecaptivus* mCherry mutant strain was detected in the PE-Texas Red Channel with a 610/20 bandpass filter.

Competitive index. Flies were microinjected with a 1:1 volume mixture of overnight cultures of wildtype (WT) *S. praecaptivus* and an *S. praecaptivus* null mutant lacking the genes *cpmA* and *cpmJ* that are known to be upregulated by QS [3]. Bacterial densities were assessed one and two weeks post introduction. Flies were surface sterilized, homogenized, and serially diluted. Dilution series were plated on non-selective media, while an equivalent amount was also plated on selective media (L-plates supplemented with gentamicin) and counted following overnight growth. Both strains are capable of growth on non-selective media, while only the *ΔcpmAJ* strain is able to grow on gentamicin. The colony count of WT *S. praecaptivus* was obtained by subtracting the colony count in the selective plate from the non-selective plate for each fly analyzed. The competitive index was calculated as follows:

$$Competitive\ index = \frac{mutant\ output/competitor\ output}{mutant\ input/competitor\ input}$$

Where mutant is the *ΔcpmAJ* strain, competitor is the *S. praecaptivus* wildtype strain, *input* is the bacteria count injected and *output* is the bacteria count recovered on plates [6].

Statistical analyses. As appropriate, statistically significant differences between treatment groups were identified with a *t*-test, non-parametric Mann-Whitney test or one-way ANOVA with Bonferroni correction for multiple comparisons using Prism version 8.1.2 (GraphPad Software, CA). Tsetse survival curves were created using the Kaplan-Meier method and statistically analyzed using the log-rank test [7] with JMP software version 10.0.1 (SAS Institute, USA). Least squares regression lines for comparing the fecundity of control and WT *S. praecaptivus* injected flies were obtained through multiple regression analysis using Prism.

References cited in Additional materials and methods

1. Dale, C. and S.C. Welburn, *The endosymbionts of tsetse flies- manipulating host-parasite interactions.* International Journal of Parasitology, 2001. **31**: p. 628-631.
2. Snyder, A.K., Adkins, K.Z. and R.V.M. Rio, *Use of the Internal Transcribed Spacer (ITS) regions to examine symbiont divergence and as a diagnostic tool for Sodalis-related bacteria.* Insects, 2011. **2**: p. 515-531.
3. Enomoto, S., et al., *Quorum Sensing Attenuates Virulence in Sodalis praecaptivus.* Cell Host Microbe, 2017. **21**(5): p. 629-636 e5.
4. Balleza, E., J.M. Kim, and P. Cluzel, *Systematic characterization of maturation time of fluorescent proteins in living cells.* Nat Methods, 2018. **15**(1): p. 47-51.
5. Kim, J.M., et al., *Stochastic transcriptional pulses orchestrate flagellar biosynthesis in.* Sci Adv, 2020. **6**(6): p. eaax0947.
6. Silver, A.C., et al., *Interaction between innate immune cells and a bacterial type III secretion system in mutualistic and pathogenic associations.* Proc Natl Acad Sci U S A, 2007. **104**(22): p. 9481-6.
7. Bewick, V., L. Cheek, and J. Ball, *Statistics review 12: survival analysis.* Crit Care, 2004. **8**(5): p. 389-94.

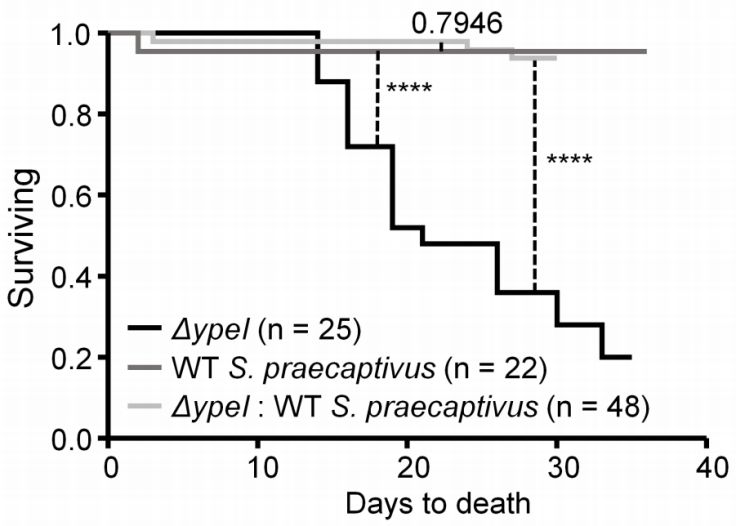

Figure 7. Rescue effect on tsetse survival upon co-injection of ΔypeI mutant with an equivalent amount of *S. praecaptivus* WT

Kaplan-Meier curves comparing survival of tsetse lines injected with WT, *S. praecaptivus* ΔypeI, and a co-injection of ΔypeI and WT. There was no significant difference in tsetse survival between WT and those tsetse receiving the coinjection (p = 0.79). **** p < 0.0001. n = number of flies.

https://doi.org/10.1371/journal.pgen.1008992.s002

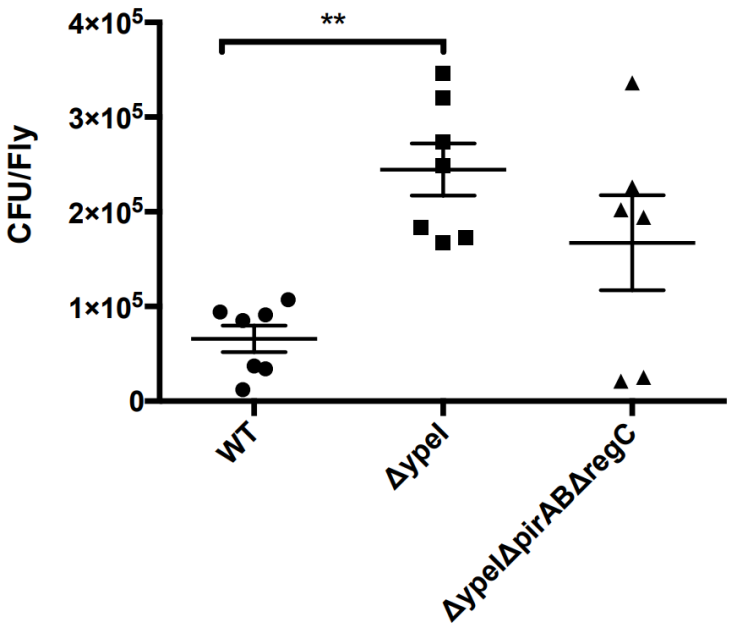

Figure 8. The density of S. praecaptivus within flies following 20 d post introduction was assessed through CFU counts after surface sterilization, homogenization, serial dilution and plating of whole flies.

Horizontal lines represent mean bacterial count, while dots represent mean bacterial count per individual. Mean bacterial density was compared (ANOVA, Tukey's multiple comparisons test **, p = 0.0022). Bars represent 1 SEM.

https://doi.org/10.1371/journal.pgen.1008992.s003

Figure 9. Symbiont infection status of progeny produced by Streptozotocin fed parental line.
Lanes correspond to; 1, wildtype whole fly DNA; 2–8 whole fly DNA of streptozotocin line. Individuals used for 2–8 were from the fourth generation of the Streptozotocin-treated parental line, which was the generation used for examining the impact of endogenous *S. glossinidius* towards *S. praecaptivus* prevalence and density.
https://doi.org/10.1371/journal.pgen.1008992.s004

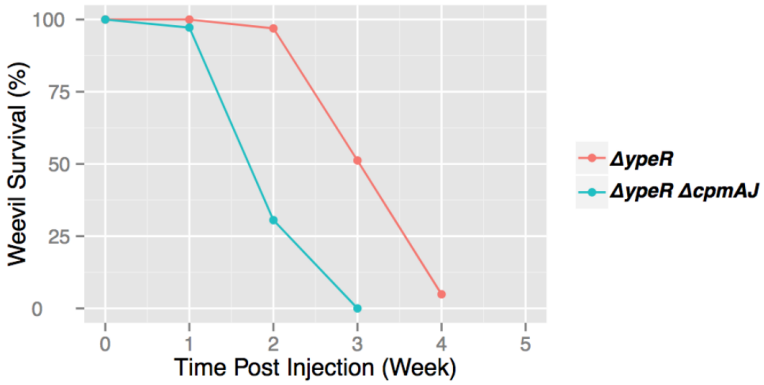

Figure 10. CpmAJ has a host protective function in weevils.
Loss of YpeR and CpmAJ synergistically accelerate weevil demise. The difference between the two survival curves was statistically significant (Logrank test: p = 9.1 e-15).
https://doi.org/10.1371/journal.pgen.1008992.s005

Table 2. S. praecaptivus strains used in this study.

Strain	Features	Mutated gene (NCBI-Protein ID): protein encoded [reference]
CD345	*S. praecaptivus ΔcpmAJ*::Gen	*cpmAJ* (AHF78165 and AHF78166): Putative carbapenem biosynthesis and resistance proteins, respectively [41].
CD348	*S. praecaptivus ΔcpmAJ*::Gen *ΔypeI*::Spc	*cpmAJ* (AHF78165 and AHF7816664): Putative carbapenem biosynthesis and resistance proteins, respectively / *ypeI* (AHF78570): Acyl-homoserine-lactone synthase [41].
CD286	*S. praecaptivus ΔypeR*::Spc	*ypeR* (AHF78569): Transcriptional activator of quorum-sensing system [41].
CD623	*S. praecaptivus Δsant_2908*::Tet	*hicA* (AHF77918): mRNA interferase; toxin-antitoxin system, strain used as WT with tetracycline resistance [this study].
CD1816	*S. praecaptivus ΔcpmAJ*::Gen *ΔypeR*::Spc	*cpmAJ* (AHF78165 and AHF78166): Putative carbapenem biosynthesis and resistance proteins, respectively /*ypeR* (AHF78569): Transcriptional activator of quorum-sensing system [this study].
CD433	*S. praecaptivus ΔypeR*::Gen *ΔyenR*::Spc	*ypeR* (AHF78569): Transcriptional activator of quorum-sensing system [this study]/*yenR* (AHF76244): Transcriptional regulator LuxR family [41].
CD384	*S. praecaptivus ΔyenR*::Spc	*yenR* (AHF76244): Transcriptional regulator of the LuxR family [41].
CD857	*S. praecaptivus ΔypeI*::Spc *ΔpirAB*::Gen *ΔregC*::Kan	*ypeI* (AHF78570): Acyl-homoserine-lactone synthase [41]/ *pirAB* (AHF77486 and AHF77486): insecticidal binary toxin complex PirAB / *regC* (AHF78621) homolog of a bacteriophage P2 transcriptional repressor RegC
CD298	*S. praecaptivus ΔypeI*::Spc	*ypeI* (AHF78570): Acyl-homoserine-lactone synthase [41]
CD14	*S. praecaptivus* wild-type	No mutation was introduced

https://doi.org/10.1371/journal.pgen.1008992.s006

Table 3. Percentage amino acid identity of *cpm* operon genes within *Sodalis*-allied insect symbionts. Bacteria identified as *Sodalis*-allied symbionts are considered. The nucleotide database was searched manually for genes annotated as involved in the carbapenem biosynthesis pathway. As no genes were identified by this method, the canonical Cpm proteins from *Photorhabdus laumondii* subspp. *laumondii* TT01 were run via NCBI tBLASTn, against the nr nucleotide database. Protein IDs used include CpmA (CAE12477), CpmB (CAE12478), CpmC (CAE12479), CpmD (CAE12480), CpmE (CAE12481), CpmF (CAE12482), CpmG (CAE12483), CpmH (CAE12485) and CpmJ (CAE12486). Homology hits were included if these occurred within the *cpm* region. * Hits to CpmG and CpmJ were hitting in identical locations in the *S. glossinidius*, *S. praecaptivus* and *Ca*. S. pierantonius genomes. As additional verification, the truncated proteins from *Sodalis glossinidius* SG0585 (NCBI-PROTEIN ID BAE73860) and SG0586 (NCBI-PROTEIN ID BAE73860) were also used in comparisons. The e-value cutoff was set to 10^{-7}.

Species	Taxonomy Id	CpmA	CpmB	CpmC	CpmD	CpmE	CpmF	CpmG*	CpmH	CpmJ*	SG0586	SG0585
Mikella endobia	1778264	-	-	-	-	-	-	-	-	-	-	-
Moranella endobia PCVAL	1234603	-	-	-	-	-	-	-	-	-	-	-
Ca. *Moranella endobia* PCIT	903503	-	-	-	-	-	-	-	-	-	-	-
Ca. *Hoaglandella endobia*	1778263	-	-	-	-	-	-	-	-	-	-	-
Ca. *Doolittlea endobia*	1778262	-	-	-	-	-	-	-	-	-	-	-
Ca. *Gullanella endobia*	1070130	-	-	-	-	-	-	-	-	-	-	-
S-endosymbiont of *Heteropsylla cubana*	134287	-	-	-	-	-	-	-	-	-	-	-
S-endosymbiont of *Ctenarytaina eucalypti*	1199245	-	-	-	-	-	-	-	-	-	-	-
Sodalis-like symbiont of *Philaneus spumarius* PSPU	1273402	-	-	-	-	-	-	-	-	-	-	-
Candidatus *Zinderia insecticola* CARI	884215	-	-	-	-	-	-	-	-	-	-	-
P-endosymbiont of *Henestaris halophilus*	1929246	-	-	-	-	-	-	-	-	-	-	-
Sodalis-like endosymbiont of *Proechinophthirus fluctus* str. SPI-1	1462730	-	-	-	-	-	-	-	-	-	-	-
Sodalis glossinidius str. "*morsitans*"	1173031	25.48	-	-	-	-	-	27.78	-	27.78	100	99
Ca. *Sodalis melophagi*	343509	-	-	-	-	-	-	-	-	-	-	-
Sodalis praecaptivus	1239307	25.49	-	-	-	-	-	26.67	-	26.67	93.33	86
Ca. *Sodalis pierantonius* str. SOPE	2342	25.49	-	-	-	-	-	32.08	-	32.08	92.38	88

https://doi.org/10.1371/journal.pgen.1008992.s007

References

1. Shin, S.C., et al., *Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling.* Science, 2011. **334**(6056): p. 670-4.
2. Storelli, G., et al., *Lactobacillus plantarum promotes Drosophila systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing.* Cell Metab, 2011. **14**(3): p. 403-14.
3. Tsuchida, T., et al., *Symbiotic bacterium modifies aphid body color.* Science, 2010. **330**(6007): p. 1102-4.
4. Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027-31.
5. Rook, G., et al., *Evolution, human-microbe interactions, and life history plasticity.* Lancet, 2017. **390**(10093): p. 521-530.
6. Kondorosi, E., P. Mergaert, and A. Kereszt, *A paradigm for endosymbiotic life: cell differentiation of Rhizobium bacteria provoked by host plant factors.* Annu Rev Microbiol, 2013. **67**: p. 611-28.
7. Nyholm, S.V. and M.J. McFall-Ngai, *The winnowing: establishing the squid-vibrio symbiosis.* Nat Rev Microbiol, 2004. **2**(8): p. 632-42.
8. Sorensen, M.E.S., et al., *The role of exploitation in the establishment of mutualistic microbial symbioses.* FEMS Microbiol Lett, 2019. **366**(12).
9. Novakova, E., V. Hypsa, and N.A. Moran, *Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution.* BMC microbiology, 2009. **9**: p. 143.
10. Werren, J.H. and D.M. Windsor, *Wolbachia infection frequencies in insects: evidence of a global equilibrium?* Proc Biol Sci, 2000. **267**(1450): p. 1277-85.
11. Gasparich, G.E., et al., *The genus Spiroplasma and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the Mycoplasma mycoides clade.* Int J Syst Evol Microbiol, 2004. **54**(Pt 3): p. 893-918.
12. Darby, A.C., et al., *Intracellular pathogens go extreme: genome evolution in the Rickettsiales.* Trends Genet, 2007. **23**(10): p. 511-20.
13. Arora, A.K. and A.E. Douglas, *Hype or opportunity? Using microbial symbionts in novel strategies for insect pest control.* J Insect Physiol, 2017. **103**: p. 10-17.
14. Rio, R.V.M., Y. Hu, and S. Aksoy, *Strategies of the home-team: Symbioses exploited for vector-borne disease control.* Trends Microbiol., 2004. **12**: p. 325-336.
15. Novakova, E. and V. Hypsa, *A new Sodalis lineage from bloodsucking fly Craterina melbae (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont Sodalis glossinidius.* FEMS Microbiol Lett, 2007. **269**(1): p. 131-5.
16. Fukatsu, T., R. Koga, W.A. Smith, K. Tanaka, N. Nikoh, K. Sasaki-Fakatus, K. Yoshizawa, C. Dale and D.H. Clayton, *Bacterial endosymbiont of the slender pigeon louse, Columbicola columbae, allied to endosymbionts fo grain weevils and tsetse flies.* Appl Environ Microbiol, 2007. **73**: p. 6660-6668.
17. Grunwald, S., M. Pilhofer, and W. Holl, *Microbial associations in guts systems of wood- and bark- inhabiting longhorned beetles [Coleoptera: Cerambyicidae].* Syst. Appl. Microbiol., 2010. **33**: p. 25-34.

18. Toju, H. and T. Fukatsu, *Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants.* Mol Ecol, 2011. **20**(4): p. 853-68.
19. Kaiwa N., T.H., Y. Kikuchi, N. Nikoh, X.Y. Meng, N. Kimura, M. Ito, and T. Fukatsu, *Primary gut symbiont and secodary, Sodalis-allied symbiont of the scutellerid stinkbug Cantao ocellatus.* Appl Environ Microbiol, 2010. **76**: p. 3486-3494.
20. Toju, H.T.H., R. Koga, N. Nikoh, X.Y. Meng, N. Kimura, and T. Fukatsu, *"Candidatus Curculioniphilus buchneri" a novel clade of bacterial endocellular symbionts from weevels of the genus Curculio.* Appl Environ Microbiol, 2010. **76**: p. 275-282.
21. Kaiwa, N., et al., *Bacterial symbionts of the giant jewel stinkbug Eucorysses grandis (Hemiptera: Scutelleridae).* Zoological science, 2011. **28**(3): p. 169-74.
22. Boyd, B.M., et al., *Two Bacterial Genera, Sodalis and Rickettsia, Associated with the Seal Louse Proechinophthirus fluctus (Phthiraptera: Anoplura).* Appl Environ Microbiol, 2016. **82**(11): p. 3185-97.
23. Santos-Garcia, D., et al., *The All-Rounder Sodalis: A New Bacteriome-Associated Endosymbiont of the Lygaeoid Bug Henestaris halophilus (Heteroptera: Henestarinae) and a Critical Examination of Its Evolution.* Genome Biol Evol, 2017. **9**(10): p. 2893-2910.
24. Rubin, B.E.R., et al., *Social behaviour in bees influences the abundance of Sodalis (Enterobacteriaceae) symbionts.* R Soc Open Sci, 2018. **5**(7): p. 180369.
25. Clayton, A.L., et al., *A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses.* PLoS Genet, 2012. **8**(11): p. e1002990.
26. Snyder, A.K., Adkins, K.Z. and R.V.M. Rio, *Use of the Internal Transcribed Spacer (ITS) regions to examine symbiont divergence and as a diagnostic tool for Sodalis-related bacteria.* Insects, 2011. **2**: p. 515-531.
27. Enomoto, S., et al., *Quorum Sensing Attenuates Virulence in Sodalis praecaptivus.* Cell Host Microbe, 2017. **21**(5): p. 629-636 e5.
28. Nealson, K.H., T. Platt, and J.W. Hastings, *Cellular control of the synthesis and activity of the bacterial luminescent system.* J Bacteriol, 1970. **104**(1): p. 313-22.
29. Fuqua, W.C., S.C. Winans, and E.P. Greenberg, *Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.* J Bacteriol, 1994. **176**(2): p. 269-75.
30. Papenfort, K. and B.L. Bassler, *Quorum sensing signal-response systems in Gram-negative bacteria.* Nat Rev Microbiol, 2016. **14**(9): p. 576-88.
31. Toh, H., 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): p. 149-56.
32. Weiss, B.L., et al., *Colonization of the tsetse fly midgut with commensal Kosakonia cowanii Zambiae inhibits trypanosome infection establishment.* PLoS Pathog, 2019. **15**(2): p. e1007470.
33. Waterfield, N., et al., *The Photorhabdus Pir toxins are similar to a developmentally regulated insect protein but show no juvenile hormone esterase activity.* FEMS Microbiol Lett, 2005. **245**(1): p. 47-52.
34. Pontes, M.H., et al., *Quorum sensing primes the oxidative stress response in the insect endosymbiont, Sodalis glossinidius.* PLoS One, 2008. **3**(10): p. e3541.

35. Dale, C.and S.C.Welburn, *The endosymbionts of tsetse flies- manipulating host-parasite interactions.* International Journal of Parasitology, 2001. **31**: p. 628-631.
36. Dale, C.and I.Maudlin, *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**: p. 267-275.
37. Demirbas-Uzel, G., et al., *Combining paratransgenesis with SIT: impact of ionizing radiation on the DNA copy number of Sodalis glossinidius in tsetse flies.* BMC Microbiol, 2018. **18**(Suppl 1): p. 160.
38. Phelps, R.J. and G.A. Vale, *Studies on populations of Glossina morsitans morsitans and G. pallidipes (Diptera: Glossinidae) in Rhodesia.* Journal of Applied Ecology, 1978. **15**(3): p. 743-760.
39. Cheng, Q. and S. Aksoy., *Tissue tropism, transmission, and expression of foreign genes in vivo in midgut symbionts of tsetse flies.* Insect Mol. Biol., 1999. **8**(1): p. 125-132.
40. Ma, W.C.and D.L.Denlinger, *Secretory discharge and microflora of milk gland in tsetse flies.* Nature, 1974. **247**: p. 301-303.
41. Balmand, S., et al., *Tissue distribution and transmission routes for the tsetse fly endosymbionts.* J Invertebr Pathol, 2013. **112 Suppl**: p. S116-22.
42. De Vooght, L., et al., *Paternal Transmission of a Secondary Symbiont during Mating in the Viviparous Tsetse Fly.* Mol Biol Evol, 2015. **32**(8): p. 1977-80.
43. Bennett, G.M. and N.A. Moran, *Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole.* Proc Natl Acad Sci U S A, 2015. **112**(33): p. 10169-76.
44. Wiegmann, B.M., et al., *Single-copy nuclear genes resolve the phylogeny of the holometabolous insects.* BMC Biol, 2009. **7**: p. 34.
45. Baker, J.E., *Development of four strains of Sitophilus oryzae (L.) (Coleoptera: Curculionidae) on barley, corn (maize), rice and wheat.* Journal of Stored Products Research, 1988. **24**(4): p. 193-198.
46. Weitz, B., *The feeding habits of Glossina.* Bull World Health Organ, 1963. **28**(5-6): p. 711-29.
47. Oakeson, K.F., et al., *Genome degeneration and adaptation in a nascent stage of symbiosis.* Genome Biol Evol, 2014. **6**(1): p. 76-93.
48. Moran, N.A. and A. Mira, *The process of genome shrinkage in the obligate symbiont Buchnera aphidicola.* Genome Biol., 2001. **2**: p. 00054.1-0054.12.
49. Chapuis, E., et al., *Virulence and pathogen multiplication: a serial passage experiment in the hypervirulent bacterial insect-pathogen Xenorhabdus nematophila.* PLoS One, 2011. **6**(1): p. e15872.
50. Chattopadhyay, A., N.B. Bhatnagar, and R. Bhatnagar, *Bacterial insecticidal toxins.* Crit Rev Microbiol, 2004. **30**(1): p. 33-54.
51. Gross, R., et al., *Immunity and symbiosis.* Mol Microbiol, 2009. **73**(5): p. 751-9.
52. Tichy, E.M., B.F. Luisi, and G.P. Salmond, *Crystal structure of the carbapenem intrinsic resistance protein CarG.* J Mol Biol, 2014. **426**(9): p. 1958-70.
53. Alizon, S., et al., *Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future.* J Evol Biol, 2009. **22**(2): p. 245-59.
54. Sachs, J.L., R.G. Skophammer, and J.U. Regus, *Evolutionary transitions in bacterial symbiosis.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108 Suppl 2**: p. 10800-7.

55. De Vooght, L., S. Van Keer, and J. Van Den Abbeele, *Towards improving tsetse fly paratransgenesis: stable colonization of Glossina morsitans morsitans with genetically modified Sodalis.* BMC Microbiol, 2018. **18**(Suppl 1): p. 165.
56. Kariithi, H.M., et al., *Enhancing vector refractoriness to trypanosome infection: achievements, challenges and perspectives.* BMC Microbiol, 2018. **18**(Suppl 1): p. 179.
57. Moloo, S.K., *An artificial feeding technique for Glossina.* Parasitology, 1971. **63**: p. 507-512.
58. Balleza, E., J.M. Kim, and P. Cluzel, *Systematic characterization of maturation time of fluorescent proteins in living cells.* Nat Methods, 2018. **15**(1): p. 47-51.
59. Kim, J.M., et al., *Stochastic transcriptional pulses orchestrate flagellar biosynthesis in.* Sci Adv, 2020. **6**(6): p. eaax0947.

CHAPTER 3: 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,#]

[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

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.

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 ancient mutualist, *Wigglesworthia*, likely established at the commencement (or soon thereafter) of tsetse species radiation with subsequent co-diversification 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, 25].

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 [26]). Although numerous microbiota contributions towards tsetse biology have been functionally characterized [27-29], 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 twenty 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 x 51 bp in length) from each sample. All reads were scored as "very good" quality (Phred scores > 30, Figure 20, Appendix B). 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* (Fig. 21, Appendix B) using Salmon [30]. 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 [31]. Non-specific 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 $\pm$ 889174, total *G. morsitans* mean reads $\pm$ SEM = 17,036,187 $\pm$ 850,383, Unpaired Student's *t*-test, p = 0.57, Fig. 22A, Appendix B) 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, Fig. 22B, Appendix B) between tsetse species libraries.

Wigglesworthia-based analyses

As *Wigglesworthia* shapes multiple aspects of tsetse physiology, immune system maturation, and metabolism [17, 27-29, 32-37], 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 $\geq$ 100 Transcripts per million (TPM) (following [38-41]). 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. 11). 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; Fig. 23A, Appendix B). 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; Fig. 23B, Appendix B). 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, $p < 0.0001$; Fig. 23C, Appendix B), supporting greater *Wigglesworthia* activity within *G. morsitans* and, although speculative, higher symbiont reliance.

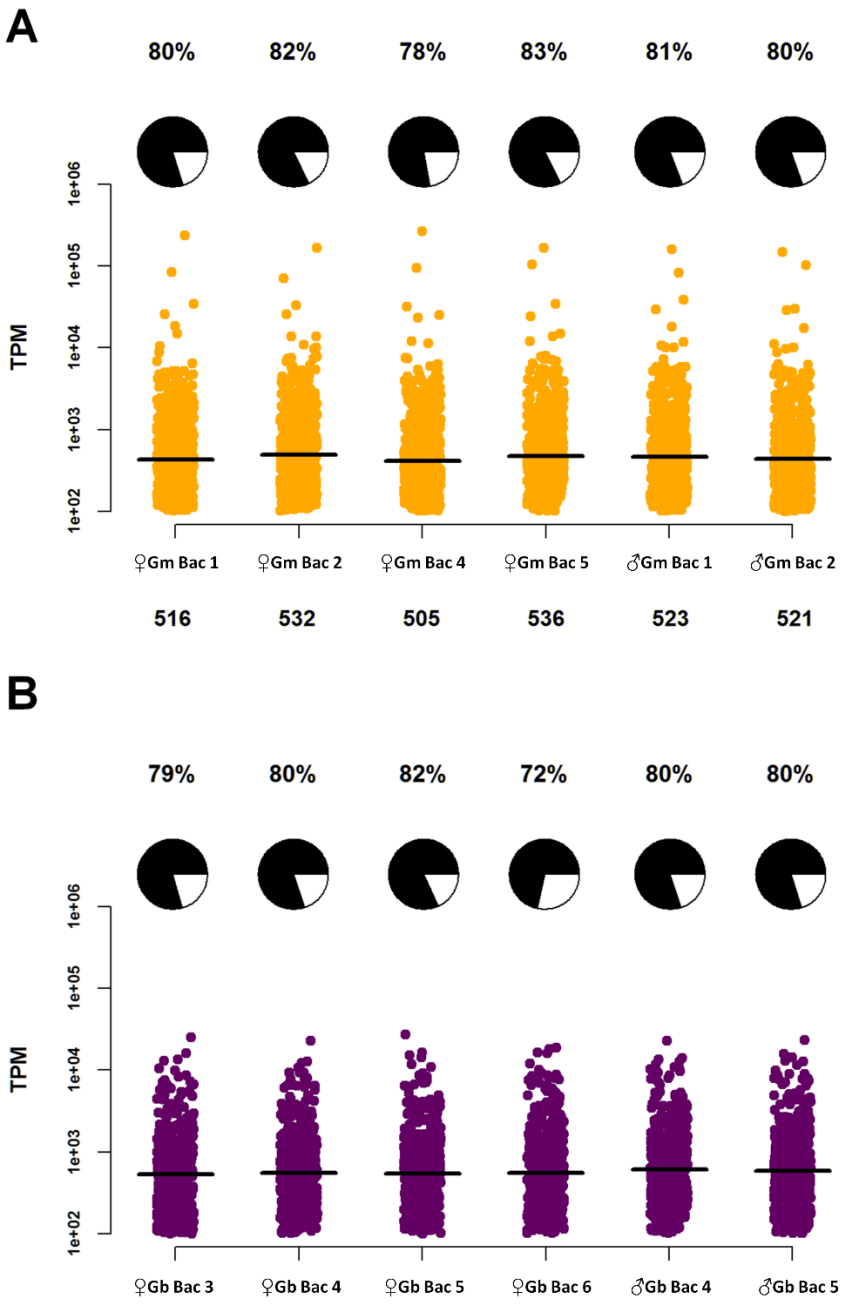

Figure 11. *Wigglesworthia* gene expression per tsetse species.
Median TPM values (for genes with TPM ≥ than 100) are shown as horizontal black lines for A) *Wigglesworthia* within teneral *G. morsitans* bacteriomes and B) *Wigglesworthia* within teneral *G. brevipalpis* bacteriomes. Dots around the 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.

Principal Component Analysis (PCA) was used to compare the global gene expression of *Wigglesworthia* between female and male libraries within a host species (Figs. 12A & 12B). 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) 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. 12C).

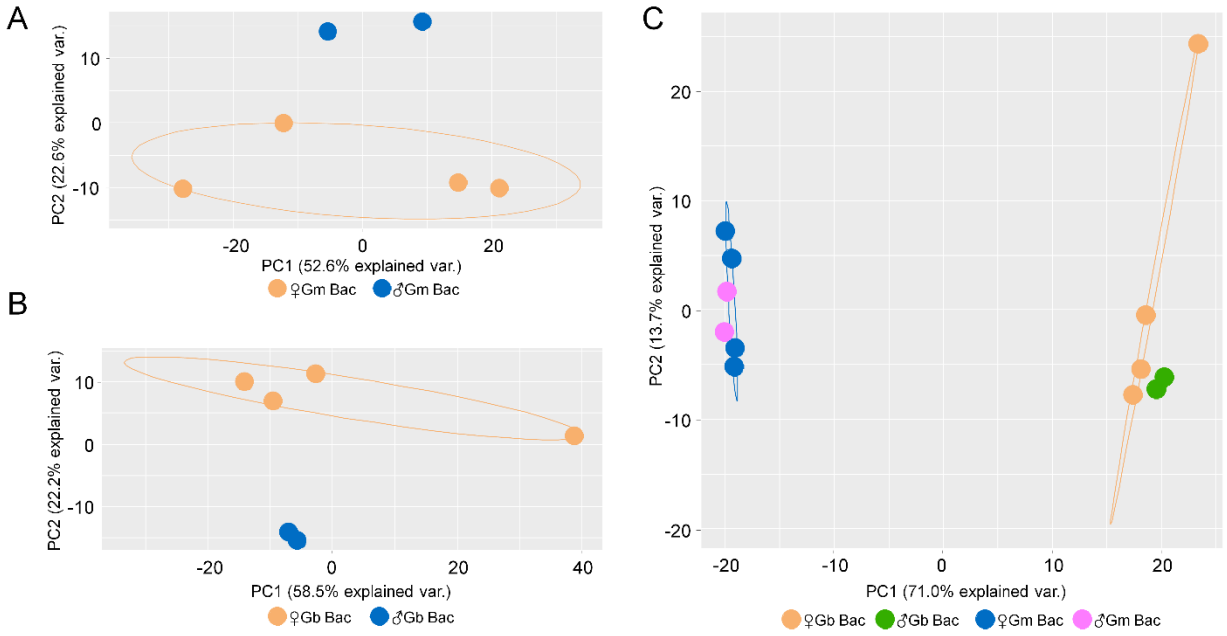

Figure 12. Principal component analysis (PCA) of *Wigglesworthia* gene expression based on TPM data.
A) PCA of *W. g. morsitans* genes sorted by sex, B) PCA of *W. g. brevipalpis* genes sorted by sex and C) PCA of 591 orthologs shared between *W. morsitans* and *W. brevipalpis* genome, sorted by sex and species. A normal data ellipse with a probability of 0.68 is shown for A-C.

Tsetse sex drives differential expression of Wigglesworthia genes

DESeq [42] 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. 13A & Table 5, Appendix B), 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 [43], 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) [44] 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 [45, 46].

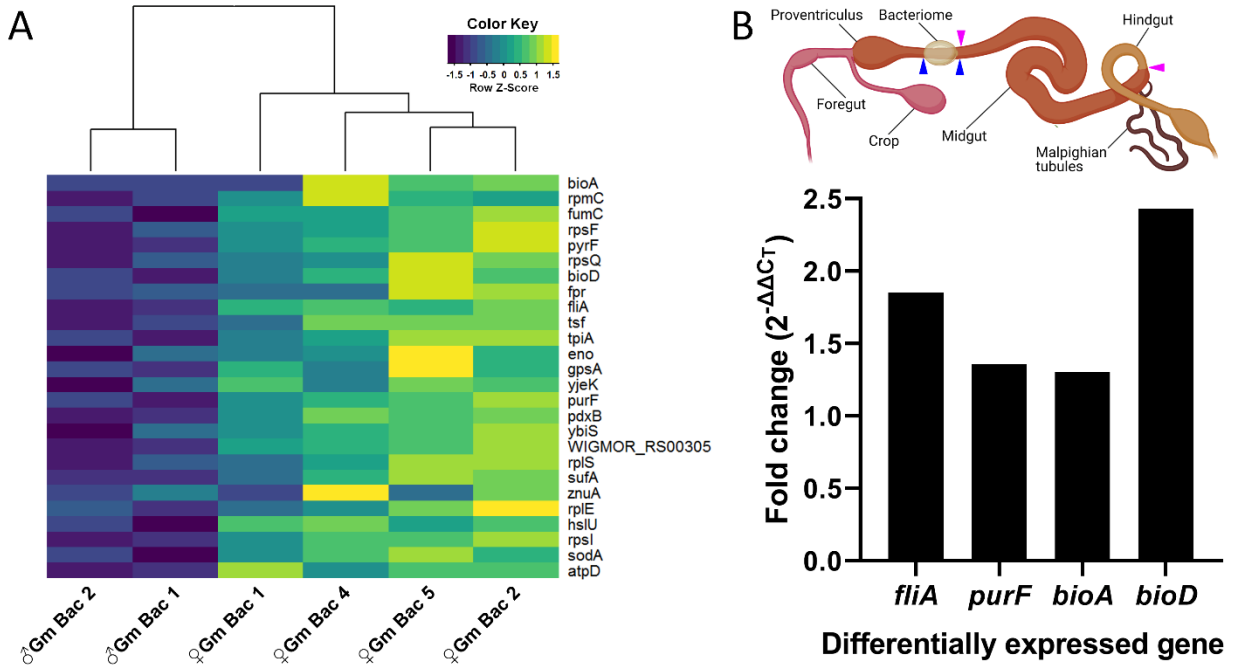

Figure 13. Differentially expressed Wigglesworthia genes between teneral males and females within G. morsitans bacteriome libraries.

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 highlighted in within the blue arrows and the midgut section, excluding the Malpighian tubules, is highlighted within pink arrows.

An elevated expression of *Wigglesworthia hslU*, a homolog of a chaperone-related protease [47, 48], was also observed within *G. morsitans* female bacteriomes. Interestingly, the oligomerization of HslU subunits is associated with the regulation of cell growth [49] and may support the significant increase in *Wigglesworthia* density during early adulthood observed in females but lacking in male tsetse flies [31]. We further confirmed the upregulation of a subset of these *Wigglesworthia* genes (*fliA*, *purF*, *bioA* and *bioD*) within female bacteriomes through qRT-PCR (Fig. 13B). 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, [50]) 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 ≥ 100, Kolmogorov-Smirnov test, $p > 0.9999$, Fig. 14A), nor for the DEGs (Kolmogorov-Smirnov test, $p = 0.9718$; Fig. 14B). 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 *Wigglesworthia*'s vitamin provisioning role following host speciation. A total of 326 orthologs (55% of 591 protein-coding genes, Fig. 14B) were identified as DEGs between the *Wigglesworthia* isolates of the two tsetse species. 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. 14B). 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.

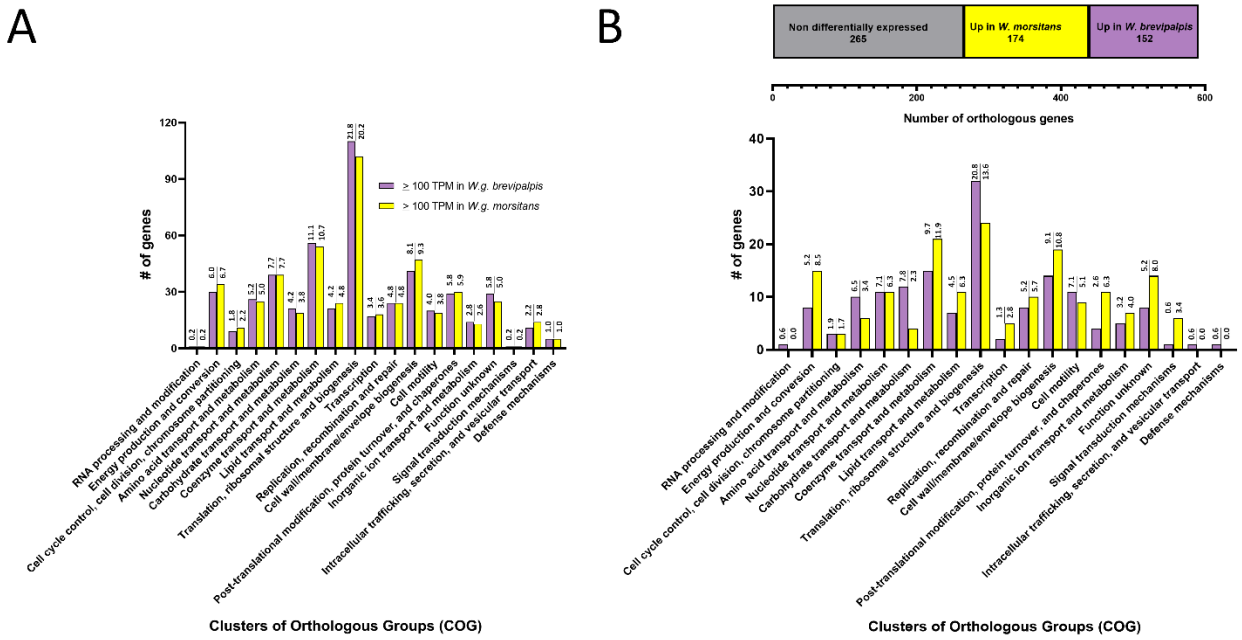

Figure 14. COG classification of highly expressed genes and DEGs between *Wigglesworthia* isolates. A) Clustering of highly expressed genes (TPM ≥ 100) into orthologous groups. Each COG shows columns for highly expressed genes in each *Wigglesworthia* isolate. The numbers on top of the bars indicate 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.

The *Wigglesworthia* genomes encode a complete flagellar apparatus [43, 51] which likely facilitate its 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. 15). The operon *flhDC* is a master regulator of flagellar genes [52] 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 [43]. 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.

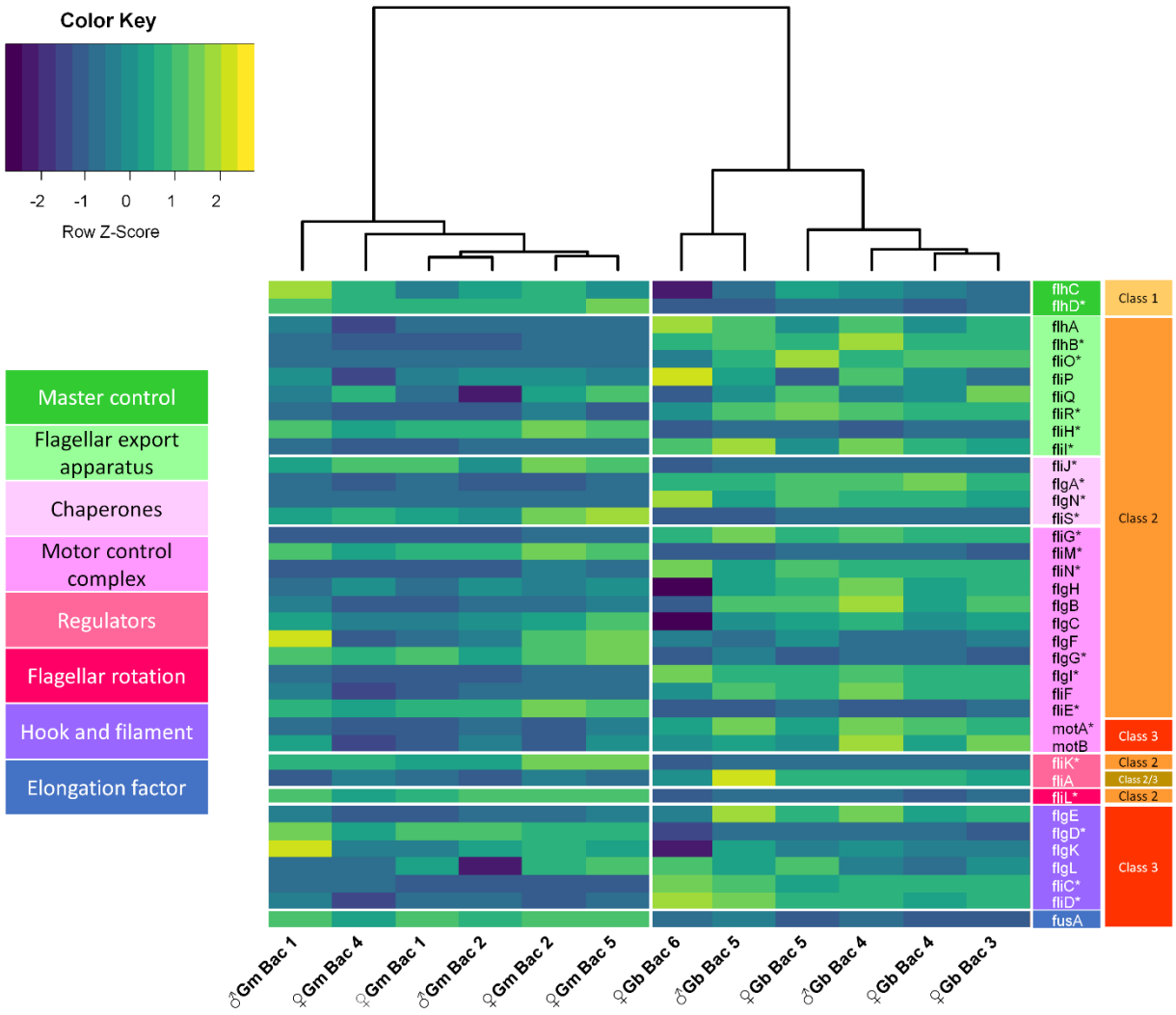

Figure 15. Distinct expression patterns of *Wigglesworthia* flagellar genes are characteristic within a tsetse host species.
A heatmap comparing the expression of flagellar genes in the *Wigglesworthia* isolates from *G. brevipalpis* and *G. morsitans* bacteriomes. Genes are vertically organized by function 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.

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 58-64% 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. 16A). If a gene was

expressed, it likely had a transcription level of < 100 TPM (i.e., >99% of 4541 genes, Fig. 16A). Only a small number of genes, constituting less than 1% of the total genes, is highly expressed (TPM $\geq$ 100). More specifically, the proportion of genes that were highly expressed is significantly larger in the *G. morsitans* isolates (26.33 $\pm$ 4.410) when compared to *G. brevipalpis* isolates (1.333 $\pm$ 0.333, Welch's test, p= 0.0291). For those genes with $\geq$100 TPM, the average TPM does not differ between the two tsetse host species (275.9 $\pm$ 26.87 in *G. morsitans* isolates vs. 208.2 $\pm$ 9.799 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. 16B), with no clustering by host sex. The analysis of read counts via IDEAmex [53] identified 235 genes to be differentially expressed between *Sodalis* isolates of the two tsetse species. 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%).

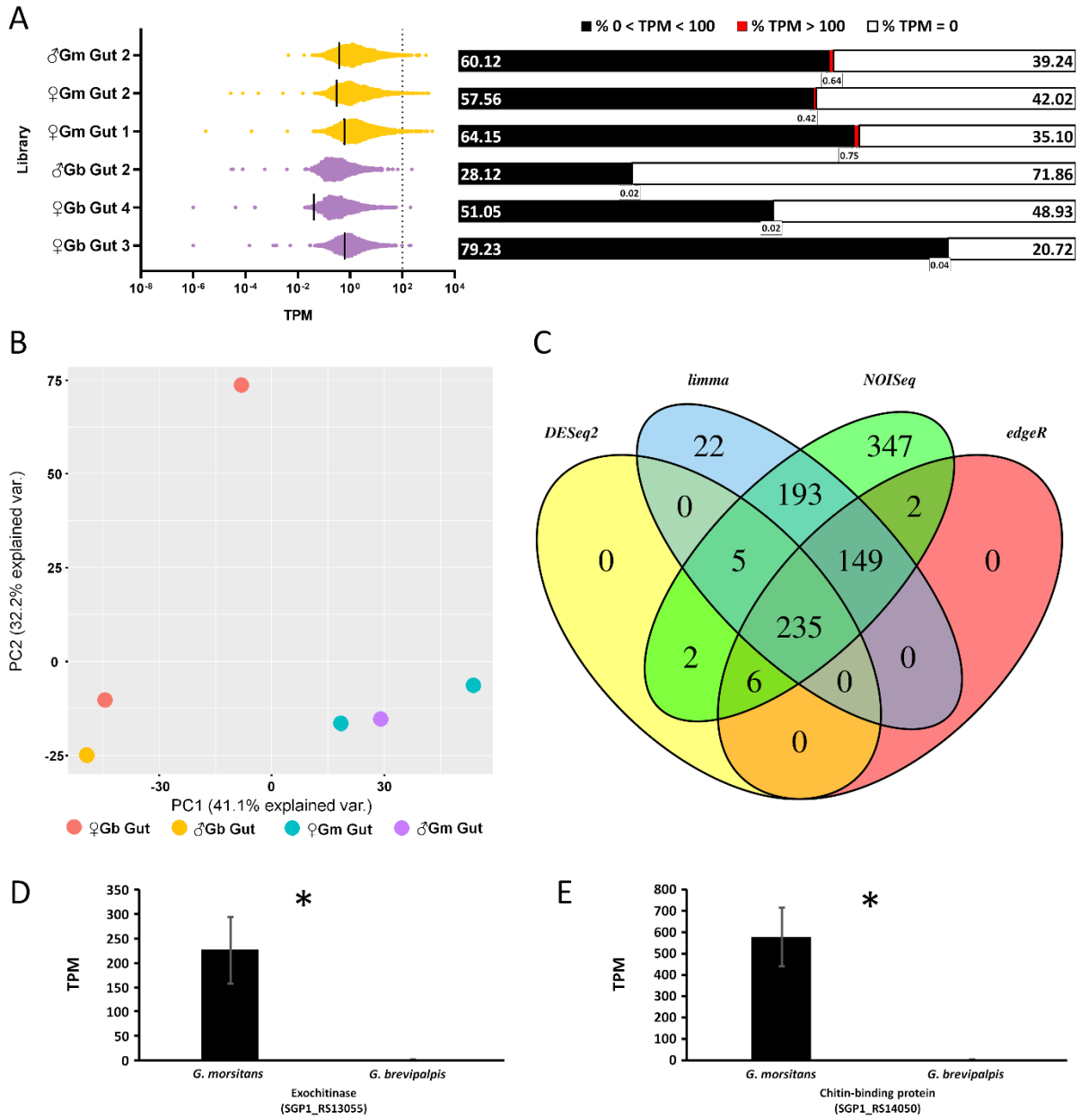

Figure 16. *Sodalis* transcriptomic profiles within teneral tsetse flies.
A) Scatterplot with TPM distribution within each tsetse midgut library; x axis is in $\log_{10}$ scale, only genes with TPM > 0 are plotted (left). Dots around the median indicate TPM values of individual genes. Bar graphs partition the percentage of the total number of genes according to their expression levels (low expression < 100 TPM, high expression ≥ 100 TPM or no expression with TPM = 0) with corresponding percentage of genes with ≥ 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 (*; $p < 0.0001$).

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 [54] coupled with the higher susceptibility of tsetse flies towards trypanosome infection as tenerals [55], *Sodalis* genes encoding chitin-associated proteins may compromise PM integrity and thereby facilitate trypanosome infection [56-58]. Two genes involved in chitin-associated activities were among the most differentially expressed between the two *Sodalis* isolates (Figs. 16D and 16E). 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 (Figs. 16D and 16E). The exochitinase gene is essential for *Sodalis* persistence within tsetse [24] as its chitinolytic activity produces *N*-acetyl-D-glucosamine (Glc*N*Ac) which is the principal carbon source for this bacterium. Additionally, the Glc*N*Ac monosaccharides inhibit anti-trypanosomal 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

Interestingly, the highly expressed genes (TPM $\geq$ 100) in the *Sodalis* transcriptome contain also a disproportionately high number of DEGs (Fig. 17). 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. 17) 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 biofilm 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 Glc*N*Ac as a 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].

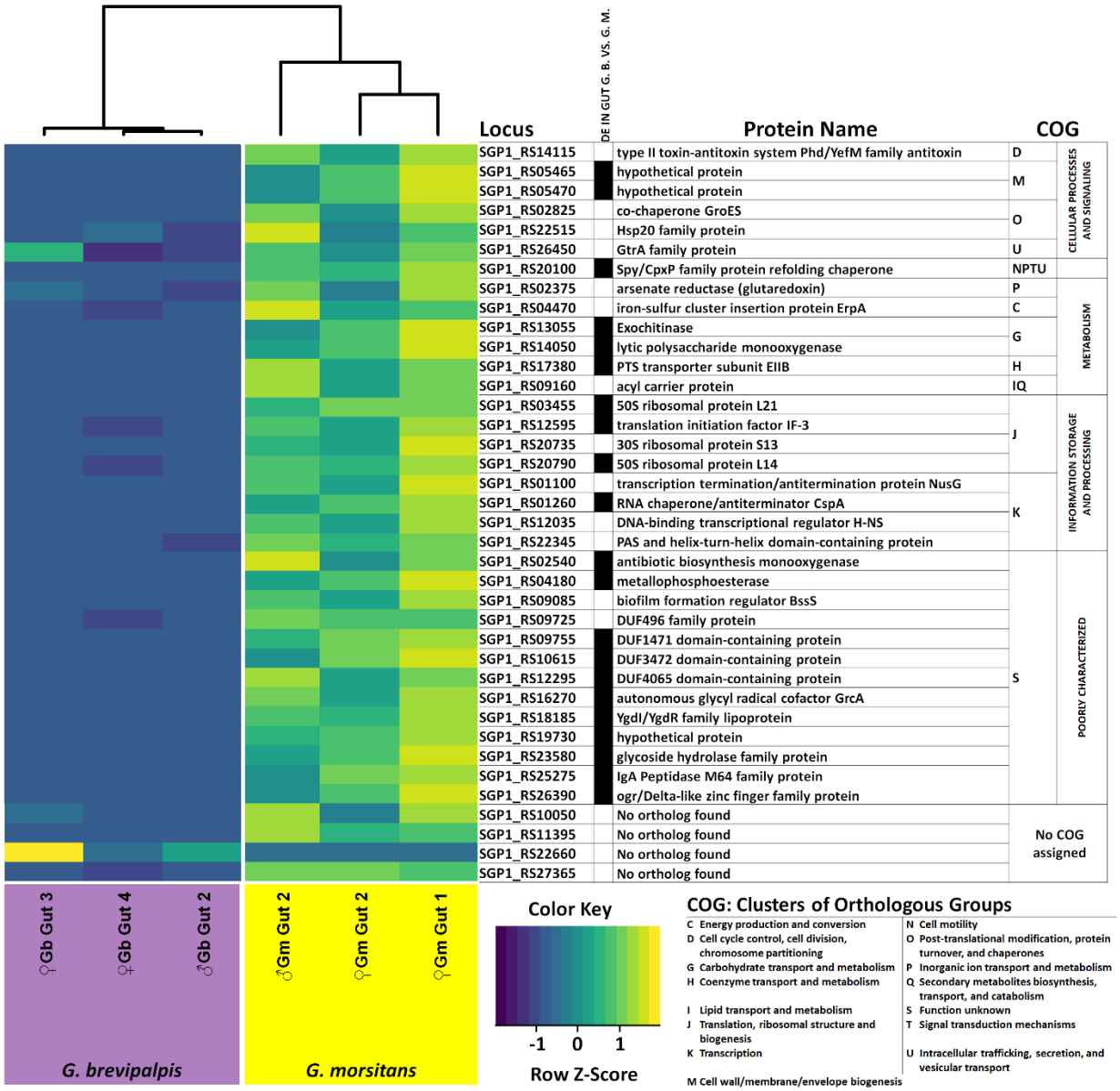

Figure 17. Highly expressed genes in the *Sodalis* transcriptome of two tsetse species.
Heatmap of *Sodalis* genes that have a high expression (TPM ≥ 100) in at least one library. The columns at the right show the locus identifier, a black square for significant **d**ifferential **e**xpression between mid**gut** 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. 18A). Only a relatively small proportion (~ 8.2%) of the total gene count is highly expressed (TPM $\geq$ 100); however, this number is significantly larger in the *G. brevipalpis* isolates [bacteriomes = 435.7 $\pm$ 24.65 (n = 6 libraries); midguts = 1197 $\pm$ 8.212 (n = 3 libraries)] when compared to *G. morsitans* [bacteriomes = 183.2 $\pm$ 10.45 (n = 6 libraries); midguts = 1007 $\pm$ 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 $\geq$ 100 TPM within midguts, the average TPM does not differ between the two tsetse host species [685.0 $\pm$ 44.69 (n = 3590 genes) in *G. brevipalpis* vs. 737.6 $\pm$ 70.59 (n = 3020 genes) in *G. morsitans* isolates, nested *t*-test, $p = 0.5158$]. However, when comparing genes with $\geq$100 TPM within bacteriome libraries, the average TPM is significantly higher for *G. brevipalpis* (overall average TPM $\pm$ SEM of 675.9 $\pm$ 52.98, n = 2614 genes) in comparison to *G. morsitans* (465.1 $\pm$ 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 for by tsetse species (PC2, Fig. 18B), with no clustering by fly sex. Differential expression analyses via IDEAmex shows 3246 DEG between *G. brevipalpis* and *G. morsitans* in bacteriome libraries (Fig. 18C) and 3134 DEG between *G. brevipalpis* and *G. morsitans* in midgut libraries (Fig. 18D).

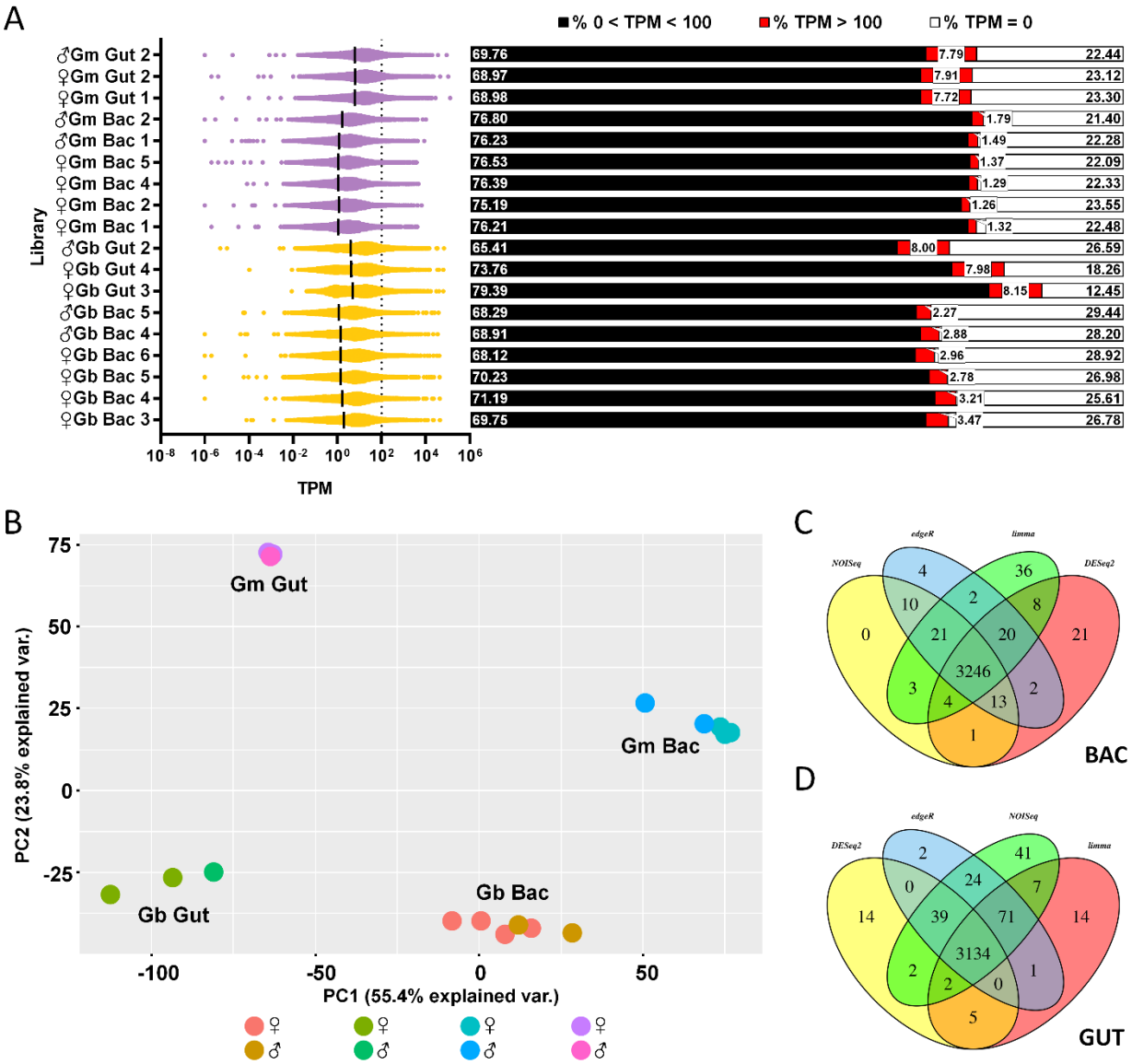

Figure 18. Transcriptomic profiles of tsetse genes within teneral tsetse flies.
A) Scatterplot with TPM distribution within each tsetse library; x axis is in $\log_{10}$ scale, only genes with TPM > 0 are plotted (left). The median is indicated with a vertical black line, TPM of 100 is indicated with 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 bottom 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.

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. 19). 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 $\pm$ 3.442 TPM in midguts vs. 15.88 $\pm$ 1.154 TPM in bacteriomes; n = 86, nested t-test, $p < 0.0001$). This immune tolerance is further 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 *Wiggleworthia* cell division, thus preventing the activation of the hostile IMD pathway and offering symbiont protection [37, 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].

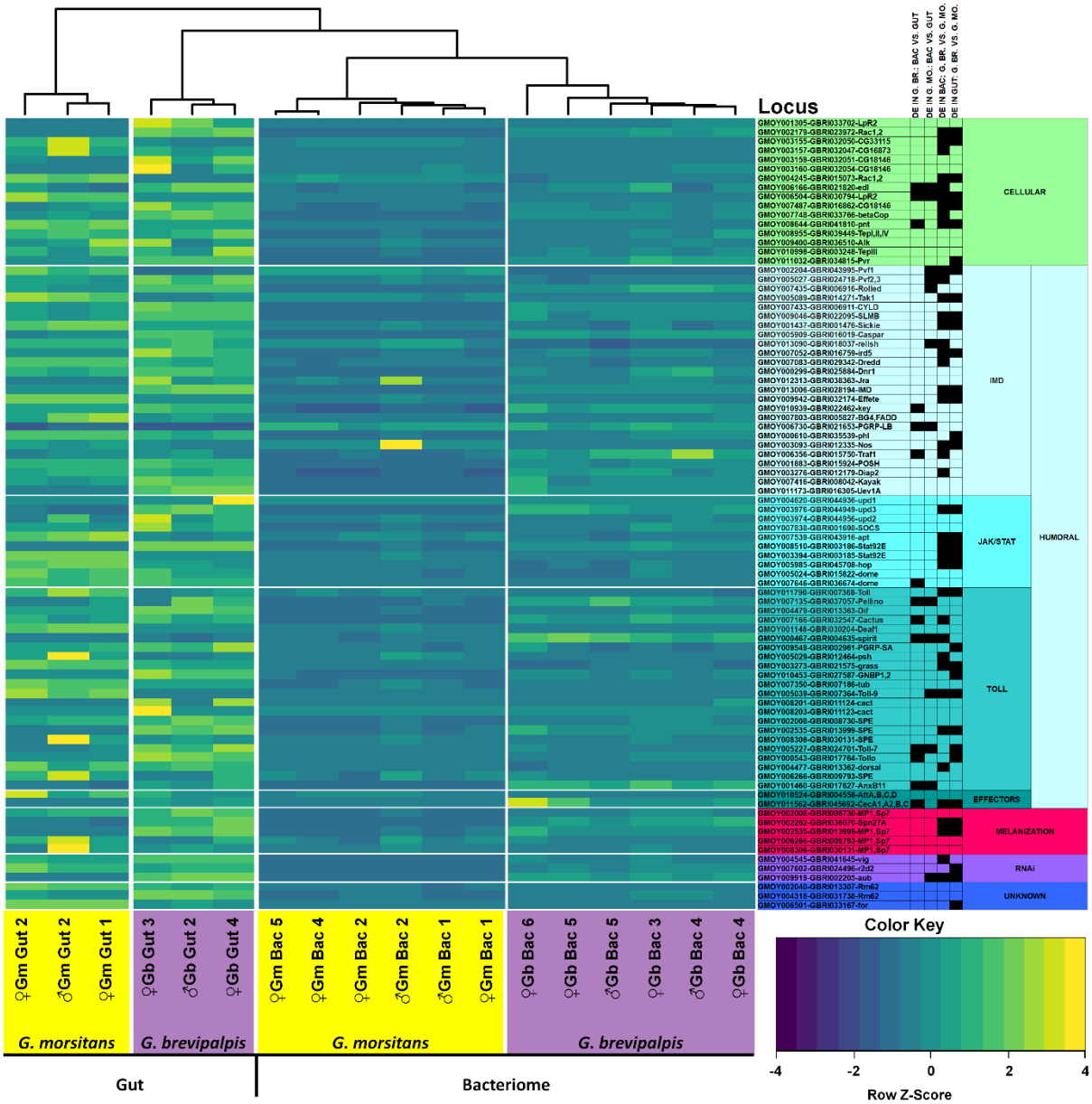

Figure 19. Expression profile of immunity genes in teneral *G. morsitans* and *G. brevipalpis*.
Heatmap with row-normalized expression levels are shown where each row represents a gene and each cell represents the relative expression level for a sample of midguts or bacteriomes in terms of Z-scores [observed 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. 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.

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 $\pm$ 1.401 TPM in JAK/STAT genes (n = 10) vs. 23.12 $\pm$ 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 coevolution 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 [26, 55, 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 [54, 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 [28, 32]. Intriguingly, between 72-83% 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 [32], 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 are 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 [52] 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 fine-tuned to its host species due to their extensive co-evolutionary 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 [34, 35], given that larvae reared in the absence of this symbiont exhibit compromised humoral and cellular immune responses to pathogens as adults [35, 36].

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, 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 [37] 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 bumblebees [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 < 24h 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 ug of total RNA and Ribo-Zero Gold Epidemiology kit (Illumina, San Diego, CA) following the manufacturer's 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 [30]. 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 [53], 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, [50]) 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 the 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 Genes

DEGs: Differentially Expressed Genes

PCA: Principal Component Analysis

PM: Peritrophic matrix

PPO: Prophenoloxidase

ROS: Reactive Oxygen Species

TPM: Transcripts per Million

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.

Availability of data and materials
Raw reads are publicly available in the Short Reads Archive (SRA) of the National Center for Biotechnology Information (Bio-project PRJNA668823; https://www.ncbi.nlm.nih.gov/bioproject/PRJNA668823).

Funding
This work was conducted with the support of the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI118789 (to R.V.M.R.). C.B. was a recipient of the Magellan Franklin Internship.

Authors' contributions
M.M.M. carried out tissue dissection, RNA isolation, sample preparation for sequencing, analyzed the data, and co-wrote the manuscript. R.V.M.R. acquired funding, designed the experiments, and co-wrote the manuscript. C.B. performed the search of orthologs between the two tsetse species. C.B., D.R. and N.S. carried out the validation of differential expression through qRT-PCR. All authors have read and approved the manuscript.

Acknowledgements

We thank Andrew Parker, at the Insect Pest Control Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Vienna, Austria, for the *G. brevipalpis* pupae. We thank the Slovak Academy of Sciences and Dr. Peter Takac for the *G. morsitans* pupae. Thanks also to the WVU Genomics Core Facility for the Illumina sequencing service, and to Dr. Aniello Infante for mapping of the raw reads. We thank Adam Pollio for contributions to the *Wigglesworthia* flagella gene analyses, and Ying Zhang for help in rearing the tsetse flies. All authors approved the final version of the manuscript.

Appendix B

Table 4. Primers used for the validation of Wigglesworthia differential expression though qRT-PCR.

Primer name	Sequence 5' – 3'	T_A (°C)	Amplicon size (bp)	Application	Target gene
GAPDH LONG For	GACAGCACCTCCACATGACA	55	596	Standard	*Wigglesworthia morsitans gapDH*
GAPDH LONG Rev	ACCCGCTTTCATTATCATACCA				
GAPDH SHORT For	TAGGATATATCGAAGAAGAT	50	131	qPCR	*Wigglesworthia morsitans gapDH*
GAPDH SHORT Rev	TTCATTATCATACCAAGATA				
fliA_short_For	CAGAGAGCGTGACTGGGTTC	55	115	qPCR	*Wigglesworthia fliA*
fliA_short_Rev	TGCTATTTGCTTTTCGGTAGCA				
fliA_long_For	TCGTCATGAAGCATTAAGATTGC	55	517	Standard	*Wigglesworthia fliA*
fliA_long_Rev	GACTTGTTCGCGTTTAGGCA				
bioA_short_For	ATGGAGATACATTTGGGGCTGT	55	126	qPCR	*Wigglesworthia bioA*
bioA_short_Rev	TCCAAGAATCGCCAAAATCACG				
bioA_long_For	CGGAATGTCATCTTGGTGGAC	55	962	Standard	*Wigglesworthia bioA*
bioA_long_Rev	CGACTCCAATAGCTCCCAGT				
purF_short_For	ACGAAGTTCAGCCATTTTACG	55	116	qPCR	*Wigglesworthia purF*
purF_short_Rev	GTGACGTCGTTGCGTTTCAA				
purF_long_For	ATGTGCGGTGTTGTTGGAAT	55	500	Standard	*Wigglesworthia purF*
purF_long_Rev	GCGCCTTGGACTCTTTGATG				
bioD_short_For	ACAGTCAATGGGTTAAGTCAGA	55	129	qPCR	*Wigglesworthia bioD*
bioD_short_Rev	TCCATCCTGCAAGTGGAACT				
bioD_long_For	AGGATGTAAATTAATAGCCTCAGGA	55	556	Standard	*Wigglesworthia bioD*
bioD_long_Rev	GAATATTTCCAATCATCGGAGCA				

Table 5. Upregulated Wigglesworthia genes within female G. morsitans isolates in comparison to male isolates.

Gene name	Gene ontology	Metabolic pathways	Specifics
bioA	Adenosylmethionine-8-amino7-oxononanoate aminotransferase	Biotin metabolism	
rpmC	50S ribosomal subunit		
fumC	fumarate hydratase class II	Metabolism	7 pathways, all metabolic
rpsF	30S ribosomal protein S6		
pyrF	orotidine-5'-phosphate decarboxylase	Metabolism	2 pathways, metabolic, pyrimidine metabolism
rpsQ	30S ribosomal subunit protein S17		
bioD	dethiobiotin synthase	Biotin metabolism	
fpr	ferredoxin-NADP reductase		Acting on iron-sulfur proteins as donors, flavodoxin
fliA	RNA polymerase sigma factor for flagellar operon		
tsf	elongation factor Tsf, protein chain elongation factor EF-Ts		Translation factor
tpiA	triosephosphate isomerase	Metabolism	9 pathways of metabolism
eno	enolase	Metabolism	Glycolysis/gluconeogenesis, 9different pathways
gpsA	glycerol-3-phosphate dehydrogenase		With NAD+ or NADP+ as acceptor
yjeK	lysine 2,3-aminomutase		Translation factor, elongation factor
purF	amidophosphoribosyltransferase	Metabolism	5 metabolic pathways
pdxB	erythronate-4-phosphate dehydrogenase	Metabolism	B6 metabolism, with NAD+ or NADP+ as acceptor
ybiS	L,D-transpeptidase		Peptidoglycan biosynthesis and degradation protein
WIGMOR_RS00305			Adjacent to *fliI*, and orthologues in other *Wigglesworthia brevipalpis, Sodalis glossinidius,* and *Sodalis praecaptivus*
rplS	50S ribosomal subunit protein L19		
sufA	Fe-S cluster assembly protein		
znuA	Periplasmic component of a high-affinity zinc uptake system	Transporter	
rplE	50S ribosomal protein L5		
hslU	molecular chaperone and ATPase component of HslUV protease		
rpsI	30S ribosomal subunit protein S9		
sodA	Superoxide dismutase, Fe-Mn family		Oxidoreductase
atpD	F1 sector of membrane-bound ATP synthase, beta subunit		Energy metabolism

Denote metabolism

Denote flagella

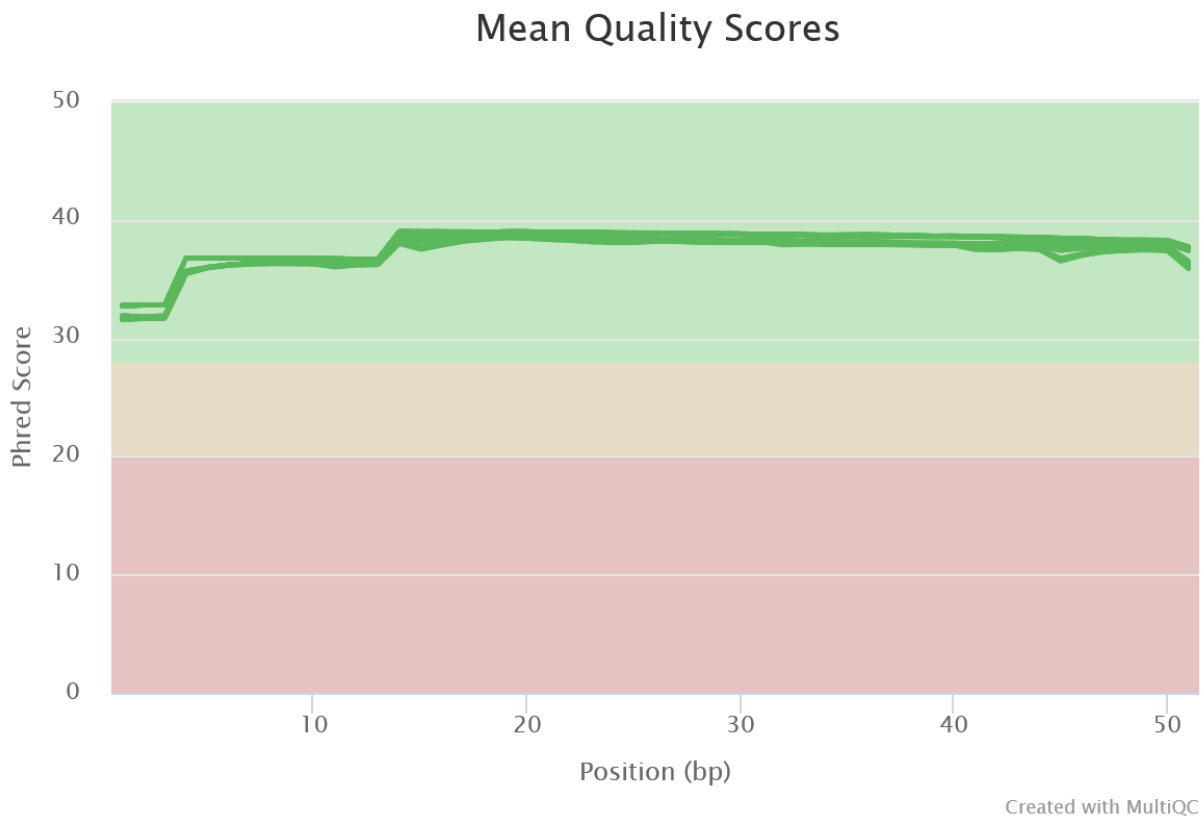

Figure 20. Mean quality scores by position of the reads.
The quality score is displayed on the y-axis, higher scores represent better qualities. The background classifies the scores into very good (in green), reasonable (in orange), and poor (in red) quality. All reads in this study were of very good quality and did not require trimming.

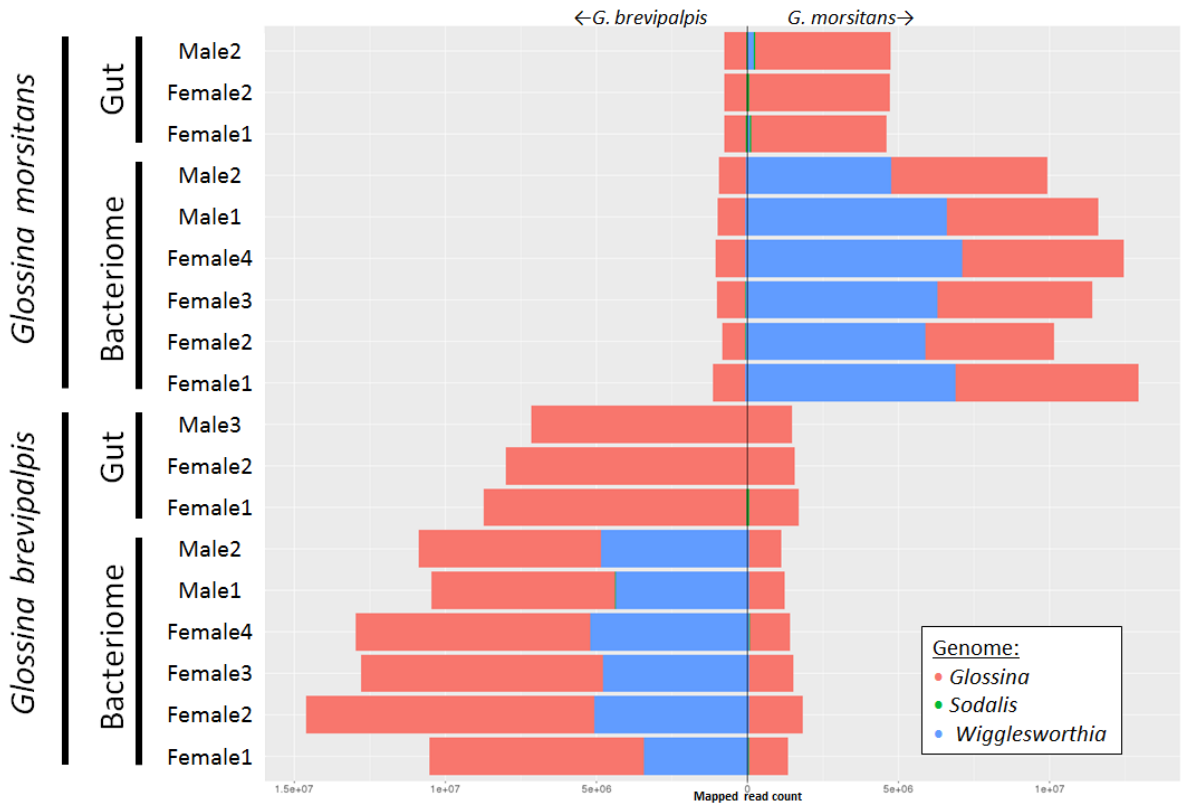

Figure 21. Read count per library.
Bar lengths represent the read count with colors indicating the genome to which the reads mapped. Reads were mapped to a construct containing either *G. brevipalpis-W. glossinidia brevipalpis-S. glossinidius* (left) or *G. morsitans-W. glossinidia morsitans-S. glossinidius* (right). The same *S. glossinidius* genome was used in both constructs, as only one *Sodalis* genome from tsetse was available at the time of analysis.

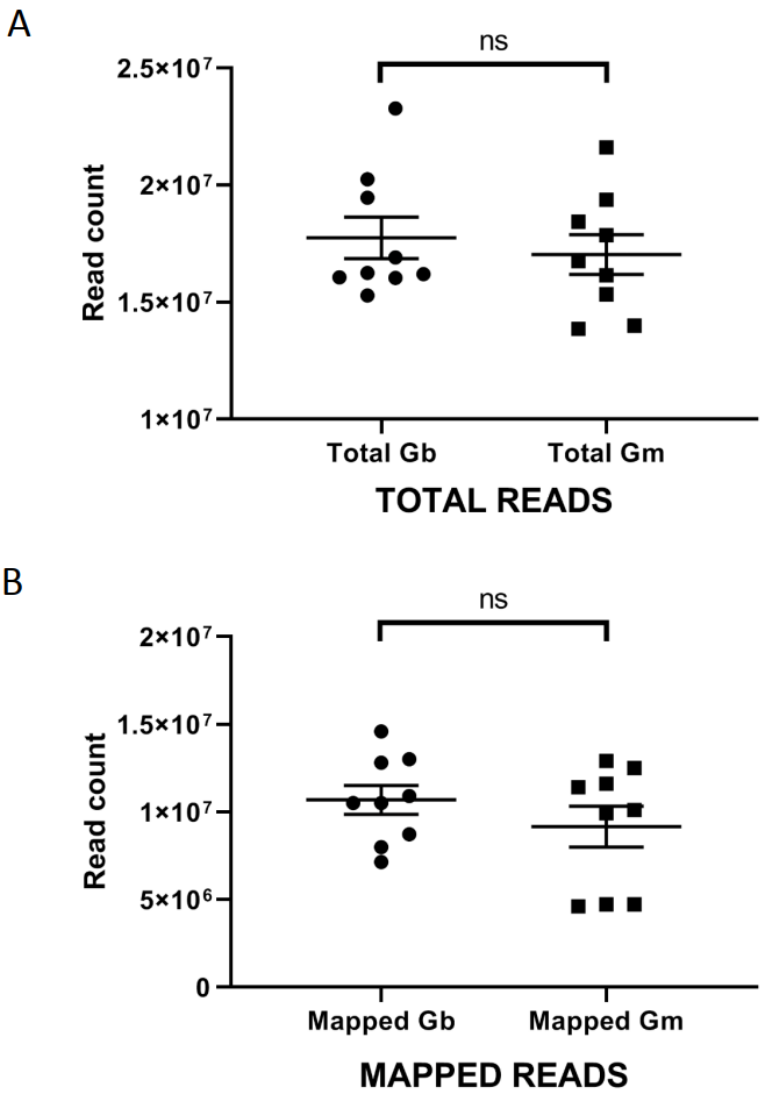

Figure 22. Comparison of total reads and mapped reads between tsetse species libraries.
A. Total reads mean ± SEM, Unpaired Student's *t*-test. B. Mapped reads mean ± SEM, Mann-Whitney test.

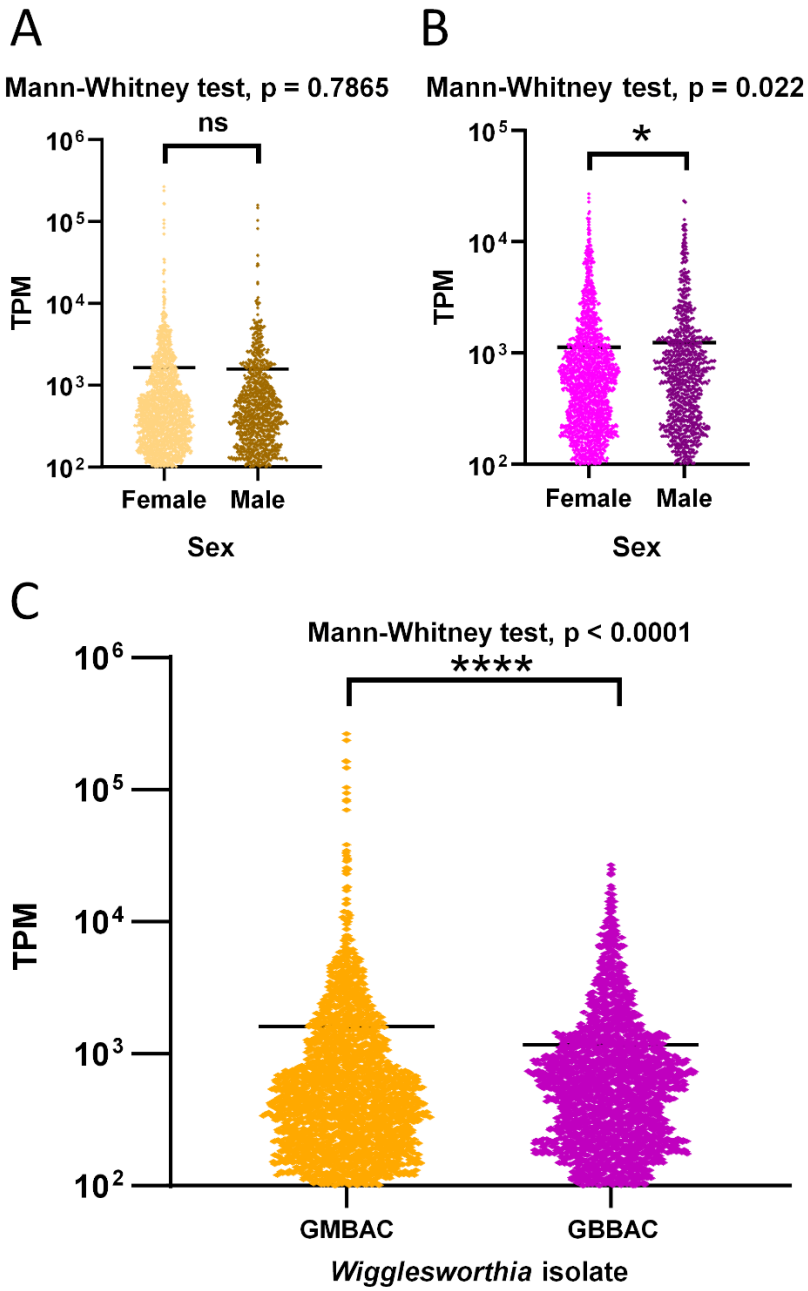

Figure 23. Within species comparison of highly expressed *Wigglesworthia* genes among two tsetse species isolates.

Highly expressed genes were defined as loci with expression levels of ≥ 100 TPM. A. Comparison of mean levels of highly expressed genes in *Wigglesworthia* expression between sexes for *G. morsitans* isolates (Mean TPM ± SEM; Mann-Whitney test). B. Comparison of mean levels of highly expressed genes in *Wigglesworthia* expression between sexes for *G. brevipalpis* isolates (Mean TPM ± SEM; Mann-Whitney test). C. Comparison of mean *Wigglesworthia* expression of highly expressed genes within bacteriome libraries between the tsetse species (Mean TPM ± SEM; Mann-Whitney test).

References

1. Simarro, P.P., et al., *Estimating and mapping the population at risk of sleeping sickness.* PLoS Negl Trop Dis, 2012. **6**(10): p. e1859.
2. Simarro, P.P., J. Jannin, and P. Cattand, *Eliminating human African trypanosomiasis: where do we stand and what comes next?* PLoS Med, 2008. **5**(2): p. e55.
3. Jordan, A.M., *Control of tsetse flies (Diptera: Glossinidae) with the aid of attractants.* J Am Mosq Control Assoc, 1995. **11**(2 Pt 2): p. 249-55.
4. Hargrove, J.W., et al., *Insecticide-treated cattle for tsetse control: the power and the problems.* Med Vet Entomol, 2000. **14**(2): p. 123-30.
5. Hargrove, J.W., S.J. Torr, and H.M. Kindness, *Insecticide-treated cattle against tsetse (Diptera: Glossinidae): what governs success?* Bull Entomol Res, 2003. **93**(3): p. 203-17.
6. Benoit, J.B., et al., *Adenotrophic viviparity in tsetse flies: potential for population control and as an insect model for lactation.* Annu Rev Entomol, 2015. **60**: p. 351-71.
7. Ma, W.C.and D.L. Denlinger, *Secretory discharge and microflora of milk gland in tsetse flies.* Nature, 1974. **247**: p. 301-303.
8. Attardo, G.M., C. Lohs, A. Heddi, U.H. Alam, S. Yildirim and S. Aksoy, *Analysis of milk gland structure and function in Glossina morsitans: Milk protein production, symbiont populations and fecundity.* J. Insect Physiol., 2008. **51**: p. 1236-1442.
9. Balmand, S., et al., *Tissue distribution and transmission routes for the tsetse fly endosymbionts.* J Invertebr Pathol, 2013. **112 Suppl**: p. S116-22.
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): p. 848-851.
11. Dale, C. and I. Maudlin, *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**: p. 267-275.
12. Aksoy, E., et al., *Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota.* Appl Environ Microbiol, 2014. **80**(14): p. 4301-12.
13. Lindh, J.M. and M.J. Lehane, *The tsetse fly Glossina fuscipes fuscipes (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts.* Antonie van Leeuwenhoek, 2011. **99**(3): p. 711-20.
14. Geiger, A., 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): p. 1364-70.
15. Chen, X.A., L. Song, and S. Aksoy, *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): p. 49-58.
16. Aksoy, S., A.A. Pourhosseini, and A. Chow, *Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to Enterobacteriaceae.* Insect Mol. Biol., 1995. **4**(1): p. 15-22.
17. Wang, J.W., et al., *Interactions between Mutualist Wigglesworthia and Tsetse Peptidoglycan Recognition Protein (Pgrp-Lb) Influence Trypanosome Transmission.* American Journal of Tropical Medicine and Hygiene, 2009. **81**(5): p. 291-291.

18. Cheng, Q.a.S.A., *Tissue tropism, transmission, and expression of foreign genes in vivo in midgut symbionts of tsetse flies.* Insect Mol. Biol., 1999. **8**(1): p. 125-132.
19. Dennis, J.W., et al., *Sodalis glossinidius prevalence and trypanosome presence in tsetse from Luambe National Park, Zambia.* Parasit Vectors, 2014. **7**: p. 378.
20. Farikou, O., F. Njiokou, J.A. Mdiba Mdiba, G.R. Njitchouang, H.N. Djeunga, T. Asonganyi, P.P. Simarro, G. Cuny, and A. Geiger *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**: p. 115-121.
21. Kame-Ngasse, G.I., et al., *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): p. 159.
22. Simo, 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**: p. 5.
23. Tsagmo Ngoune, J.M., 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): p. 151.
24. Dale, C. and S.C. Welburn, *The endosymbionts of tsetse flies: manipulating host-parasite interactions.* Int J Parasitol, 2001. **31**(5-6): p. 628-31.
25. Trappeniers, K., et al., *The Tsetse Fly Displays an Attenuated Immune Response to Its Secondary Symbiont,.* Front Microbiol, 2019. **10**: p. 1650.
26. Haines, L.R., *Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection.* Front Cell Infect Microbiol, 2013. **3**: p. 84.
27. Snyder, A.K., C. McLain, and R.V. Rio, *The tsetse fly obligate mutualist Wigglesworthia morsitans alters gene expression and population density via exogenous nutrient provisioning.* Appl Environ Microbiol, 2012. **78**(21): p. 7792-7.
28. Michalkova, V., et al., *Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies.* Appl Environ Microbiol, 2014. **80**(18): p. 5844-53.
29. Snyder, A.K. and R.V. Rio, *"Wigglesworthia morsitans" Folate (Vitamin B9) Biosynthesis Contributes to Tsetse Host Fitness.* Appl Environ Microbiol, 2015. **81**(16): p. 5375-86.
30. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017. **14**(4): p. 417-419.
31. Rio, R.V., et al., *Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora.* Proc Biol Sci, 2006. **273**(1588): p. 805-14.
32. Rio, R.V.M., et al., *Mutualist-Provisioned Resources Impact Vector Competency.* mBio, 2019. **10**(3).
33. Benoit, J.B., et al., *Symbiont-induced odorant binding proteins mediate insect host hematopoiesis.* Elife, 2017. **6**.
34. Weiss, B.L., J. Wang, and S. Aksoy, *Tsetse immune system maturation requires the presence of obligate symbionts in larvae.* PLoS Biology, 2011. **9**(5): p. e1000619.
35. Weiss, B.L., M. Maltz, and S. Aksoy, *Obligate symbionts activate immune system development in the tsetse fly.* Journal of immunology, 2012. **188**(7): p. 3395-403.

36. Weiss, B.L., et al., *Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers.* PLoS Pathog, 2013. **9**(4): p. e1003318.
37. Wang, J.W. and S. Aksoy, *PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(26): p. 10552-10557.
38. Yang, C.P., et al., *Transcriptomes of lineage-specific Drosophila neuroblasts profiled by genetic targeting and robotic sorting.* Development, 2016. **143**(3): p. 411-21.
39. Yin, W., Y. Song, and X. Chang, *Single-cell RNA-Seq analysis identifies a noncoding.* J Biol Chem, 2019. **294**(1): p. 290-298.
40. Macrander, J., et al., *Venomix: a simple bioinformatic pipeline for identifying and characterizing toxin gene candidates from transcriptomic data.* PeerJ, 2018. **6**: p. e5361.
41. John, C.R., et al., *Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses.* Plant Physiol, 2014. **165**(1): p. 62-75.
42. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
43. Rio, R.V., et al., *Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont Wigglesworthia.* MBio, 2012. **3**(1).
44. Ibuki, T., et al., *Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases.* Nat Struct Mol Biol, 2011. **18**(3): p. 277-82.
45. Tso, J.Y., et al., *Nucleotide sequence of Escherichia coli purF and deduced amino acid sequence of glutamine phosphoribosylpyrophosphate amidotransferase.* J Biol Chem, 1982. **257**(7): p. 3525-31.
46. Sampei, G. and K. Mizobuchi, *Nucleotide sequence of the Escherichia coli purF gene encoding amidophosphoribosyltransferase for de novo purine nucleotide synthesis.* Nucleic Acids Res, 1988. **16**(17): p. 8717.
47. Missiakas, D., et al., *Identification and characterization of HsIV HsIU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in Escherichia coli.* EMBO J, 1996. **15**(24): p. 6899-909.
48. Rohrwild, M., 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): p. 5808-13.
49. Seong, I.S., et al., *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): p. 224-9.
50. Tatusov, R.L., E.V. Koonin, and D.J. Lipman, *A genomic perspective on protein families.* Science, 1997. **278**(5338): p. 631-7.
51. Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy, *Genome sequence of the endocellular obligate symbiont of tsetse, Wigglesworthia glossinidia.* Nat. Gen., 2002. **32**(2): p. 402-407.
52. Liu, X. and P. Matsumura, *The FlhD/FlhC complex, a transcriptional activator of the Escherichia coli flagellar class II operons.* J Bacteriol, 1994. **176**(23): p. 7345-51.
53. Jiménez-Jacinto, V., A. Sanchez-Flores, and L. Vega-Alvarado, *Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis.* Front Genet, 2019. **10**: p. 279.

54. Weiss, B.L., et al., *The peritrophic matrix mediates differential infection outcomes in the tsetse fly gut following challenge with commensal, pathogenic, and parasitic microbes.* Journal of immunology, 2014. **193**(2): p. 773-82.
55. Van Hoof, L., C. Henrard, and E. Peel, *Influences modificatrices de la transmissibilite cyclique du Trypanosoma gambiense par Glossina palpalis.* Annales Societe Belge Med. Trop., 1937(17): p. 249-272.
56. Maudlin, I. and D.S. Ellis, *Association between intracellular rickettsial-like infections of midgut cells and susceptibility to trypanosome infection in Glossina spp.* Z Parasitenkd, 1985. **71**(5): p. 683-7.
57. Welburn, S.C., et al., *Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies.* Parasitology, 1993. **107 (Pt 2)**: p. 141-5.
58. Rose, C., et al., *An investigation into the protein composition of the teneral Glossina morsitans morsitans peritrophic matrix.* PLoS Negl Trop Dis, 2014. **8**(4): p. e2691.
59. Ibrahim, E.A., G.A. Ingram, and D.H. Molyneux, *Haemagglutinins and parasite agglutinins in haemolymph and gut of Glossina.* Tropenmed Parasitol, 1984. **35**(3): p. 151-6.
60. Maudlin, I. and S.C. Welburn, *Lectin mediated establishment of midgut infections of Trypanosoma congolense and Trypanosoma brucei in Glossina morsitans.* Trop Med Parasitol, 1987. **38**(3): p. 167-70.
61. Xu, J., et al., *Identification of Three Type II Toxin-Antitoxin Systems in.* Toxins (Basel), 2018. **10**(11).
62. Domka, J., J. Lee, and T.K. Wood, *YliH (BssR) and YceP (BssS) regulate Escherichia coli K-12 biofilm formation by influencing cell signaling.* Appl Environ Microbiol, 2006. **72**(4): p. 2449-59.
63. Courtade, G. and F.L. Aachmann, *Chitin-Active Lytic Polysaccharide Monooxygenases.* Adv Exp Med Biol, 2019. **1142**: p. 115-129.
64. Eijsink, V., I. Hoell, and G. Vaaje-Kolstada, *Structure and function of enzymes acting on chitin and chitosan.* Biotechnol Genet Eng Rev, 2010. **27**: p. 331-66.
65. Arsène, F., T. Tomoyasu, and B. Bukau, *The heat shock response of Escherichia coli.* Int J Food Microbiol, 2000. **55**(1-3): p. 3-9.
66. Guest, R.L. and T.L. Raivio, *Role of the Gram-Negative Envelope Stress Response in the Presence of Antimicrobial Agents.* Trends Microbiol, 2016. **24**(5): p. 377-390.
67. Liu, J., et al., *ATPase activity of GroEL is dependent on GroES and it is response for environmental stress in Riemerella anatipestifer.* Microb Pathog, 2018. **121**: p. 51-58.
68. Susin, M.F., et al., *GroES/GroEL and DnaK/DnaJ have distinct roles in stress responses and during cell cycle progression in Caulobacter crescentus.* J Bacteriol, 2006. **188**(23): p. 8044-53.
69. Kupper, M., et al., *Versatile roles of the chaperonin GroEL in microorganism-insect interactions.* FEMS Microbiol Lett, 2014. **353**(1): p. 1-10.
70. Stoll, S., H. Feldhaar, and R. Gross, *Transcriptional profiling of the endosymbiont Blochmannia floridanus during different developmental stages of its holometabolous ant host.* Environ Microbiol, 2009. **11**(4): p. 877-88.
71. Medina Munoz, M., et al., *Into the Wild: Parallel Transcriptomics of the Tsetse-Wigglesworthia Mutualism within Kenyan Populations.* Genome Biol Evol, 2017. **9**(9): p. 2276-2291.

72. Fares, M.A., et al., *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): p. 1162-70.
73. Hu, C., 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): p. e192.
74. Hao, Z., et al., *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): p. 12648-53.
75. Hao, Z., I. Kasumba, and S. Aksoy, *Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidae).* Insect Biochem. Mol. Biol., 2003. **33**: p. 1155-1164.
76. MacLeod, E.T., et al., *Antioxidants promote establishment of trypanosome infections in tsetse.* Parasitology, 2007. **134**: p. 827-831.
77. Vigneron, A., 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): p. e1006972.
78. Eleftherianos, I., et al., *Endosymbiotic bacteria in insects: guardians of the immune system?* Front Physiol, 2013. **4**: p. 46.
79. Medzhitov, R., D.S. Schneider, and M.P. Soares, *Disease tolerance as a defense strategy.* Science, 2012. **335**(6071): p. 936-41.
80. Obbard, D.J., et al., *Quantifying adaptive evolution in the Drosophila immune system.* PLoS Genet, 2009. **5**(10): p. e1000698.
81. Matetovici, I., G. Caljon, and J. Van Den Abbeele, *Tsetse fly tolerance to T. brucei infection: transcriptome analysis of trypanosome-associated changes in the tsetse fly salivary gland.* BMC Genomics, 2016. **17**(1): p. 971.
82. Wang , J., Y. Wu, G. Yang and S. Aksoy, *Interactions between mutualist Wigglesworthia and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypansome transmission.* Proc. Natl. Acad. Sci. U S A, 2009. **106**: p. 12134-12138.
83. Soukup, S.F., J. Culi, and D. Gubb, *Uptake of the necrotic serpin in Drosophila melanogaster via the lipophorin receptor-1.* PLoS Genet, 2009. **5**(6): p. e1000532.
84. Tsuzuki, S., et al., *Switching between humoral and cellular immune responses in Drosophila is guided by the cytokine GBP.* Nat Commun, 2014. **5**: p. 4628.
85. Hedengren, M., K. Borge, and D. Hultmark, *Expression and evolution of the Drosophila attacin/diptericin gene family.* Biochem Biophys Res Commun, 2000. **279**(2): p. 574-81.
86. Hultmark, D., et al., *Insect immunity. Attacins, a family of antibacterial proteins from Hyalophora cecropia.* EMBO J, 1983. **2**(4): p. 571-6.
87. Imler, J.L. and P. Bulet, *Antimicrobial peptides in Drosophila: structures, activities and gene regulation.* Chem Immunol Allergy, 2005. **86**: p. 1-21.
88. Lye, S.H. and S. Chtarbanova, *as a Model to Study Brain Innate Immunity in Health and Disease.* Int J Mol Sci, 2018. **19**(12).
89. Meister, S., A.C. Koutsos, and G.K. Christophides, *The Plasmodium parasite--a 'new' challenge for insect innate immunity.* Int J Parasitol, 2004. **34**(13-14): p. 1473-82.
90. Jo, Y.H., et al., *Regulation of the expression of nine antimicrobial peptide genes by TmIMD confers resistance against Gram-negative bacteria.* Sci Rep, 2019. **9**(1): p. 10138.

91. Onfelt Tingvall, T., E. Roos, and Y. Engström, *The imd gene is required for local Cecropin expression in Drosophila barrier epithelia.* EMBO Rep, 2001. **2**(3): p. 239-43.
92. Durvasula, R.V., et al., *Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria.* Proc Natl Acad Sci U S A, 1997. **94**(7): p. 3274-8.
93. Fieck, A., et al., *Trypanosoma cruzi: synergistic cytotoxicity of multiple amphipathic anti-microbial peptides to T. cruzi and potential bacterial hosts.* Exp Parasitol, 2010. **125**(4): p. 342-7.
94. Chen, L., et al., *Innate immune signaling in.* J Biol Chem, 2017. **292**(21): p. 8738-8749.
95. Kambris, Z., et al., *Drosophila immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for Toll activation.* Curr Biol, 2006. **16**(8): p. 808-13.
96. Ji, S., et al., *Cell-surface localization of Pellino antagonizes Toll-mediated innate immune signalling by controlling MyD88 turnover in Drosophila.* Nat Commun, 2014. **5**: p. 3458.
97. Germani, F., et al., *The Toll pathway inhibits tissue growth and regulates cell fitness in an infection-dependent manner.* Elife, 2018. **7**.
98. Liu, B., et al., *Toll Receptor-Mediated Hippo Signaling Controls Innate Immunity in Drosophila.* Cell, 2016. **164**(3): p. 406-19.
99. Weyd, H., et al., *Annexin A1 on the surface of early apoptotic cells suppresses CD8+ T cell immunity.* PLoS One, 2013. **8**(4): p. e62449.
100. Bollinger, A.L., et al., *Annexin V expression on CD4.* Immunology, 2020. **159**(2): p. 205-220.
101. Yang, H., et al., *JAK/STAT signaling in Drosophila muscles controls the cellular immune response against parasitoid infection.* EMBO Rep, 2015. **16**(12): p. 1664-72.
102. Agaisse, H., et al., *Signaling role of hemocytes in Drosophila JAK/STAT-dependent response to septic injury.* Dev Cell, 2003. **5**(3): p. 441-50.
103. Starz-Gaiano, M., et al., *Feedback inhibition of Jak/STAT signaling by apontic is required to limit an invasive cell population.* Dev Cell, 2008. **14**(5): p. 726-38.
104. Tang, H., et al., *Two proteases defining a melanization cascade in the immune system of Drosophila.* J Biol Chem, 2006. **281**(38): p. 28097-104.
105. Castillejo-López, C. and U. Häcker, *The serine protease Sp7 is expressed in blood cells and regulates the melanization reaction in Drosophila.* Biochem Biophys Res Commun, 2005. **338**(2): p. 1075-82.
106. De Gregorio, E., et al., *An immune-responsive Serpin regulates the melanization cascade in Drosophila.* Dev Cell, 2002. **3**(4): p. 581-92.
107. Distelmans, W., et al., *The susceptibility of Glossina palpalis palpalis at different ages to infection with Trypanosoma congolense.* Ann Soc Belg Med Trop, 1982. **62**(1): p. 41-7.
108. Otieno, L.H., et al., *Some observations on factors associated with the development of Trypanosoma brucei brucei infections in Glossina morsitans morsitans.* Acta Trop, 1983. **40**(2): p. 113-20.
109. Welburn, S.C. and I. Maudlin, *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): p. 529-36.
110. Lehane, M.J. and A.R. Msangi, *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): p. 495-501.

111. Maudlin, I. and S.C. Welburn, *The role of lectins and trypanosome genotype in the maturation of midgut infections in Glossina morsitans.* Trop Med Parasitol, 1988. **39**(1): p. 56-8.
112. Moloo, S.K., C.L. Sabwa, and J.M. Kabata, *Vector competence of Glossina pallidipes and G. morsitans centralis for Trypanosoma vivax, T. congolense and T. b. brucei.* Acta Trop, 1992. **51**(3-4): p. 271-80.
113. Motloang, M., et al., *Vector competence of Glossina austeni and Glossina brevipalpis for Trypanosoma congolense in KwaZulu-Natal, South Africa.* Onderstepoort J Vet Res, 2012. **79**(1): p. E1-6.
114. Leak, S.G.A., *Tsetse biology and ecology, their role in the epidemiology and control of trypanosomes*. 1999, New York, NY: CABI publishing.
115. Snyder, A.K., J.W. Deberry, L. Runyen-Janecky and R.V.M. Rio, *Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts.* Proc Biol Sci, 2010. 7;277(1692):2389-97.
116. Klein, C.C., 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): p. e79786.
117. Opperdoes, F.R., et al., *Comparative Metabolism of Free-living Bodo saltans and Parasitic Trypanosomatids.* J Eukaryot Microbiol, 2016. **63**(5): p. 657-78.
118. Rose, C., et al., *Trypanosoma brucei colonizes the tsetse gut via an immature peritrophic matrix in the proventriculus.* Nat Microbiol, 2020.
119. Guillén, D., S. Sánchez, and R. Rodríguez-Sanoja, *Carbohydrate-binding domains: multiplicity of biological roles.* Appl Microbiol Biotechnol, 2010. **85**(5): p. 1241-9.
120. Frederiksen, R.F., et al., *Bacterial chitinases and chitin-binding proteins as virulence factors.* Microbiology (Reading), 2013. **159**(Pt 5): p. 833-847.
121. Rossi, E., M. Paroni, and P. Landini, *Biofilm and motility in response to environmental and host-related signals in Gram negative opportunistic pathogens.* J Appl Microbiol, 2018.
122. Wang, S., et al., *Structure of the Escherichia coli FlhDC complex, a prokaryotic heteromeric regulator of transcription.* J Mol Biol, 2006. **355**(4): p. 798-808.
123. Macnab, R.M., *Genetics and biogenesis of bacterial flagella.* Annu Rev Genet, 1992. **26**: p. 131-58.
124. Chevance, F.F. and K.T. Hughes, *Coordinating assembly of a bacterial macromolecular machine.* Nat Rev Microbiol, 2008. **6**(6): p. 455-65.
125. Claret, L. and C. Hughes, *Rapid turnover of FlhD and FlhC, the flagellar regulon transcriptional activator proteins, during Proteus swarming.* J Bacteriol, 2000. **182**(3): p. 833-6.
126. Wei, B.L., et al., *Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli.* Mol Microbiol, 2001. **40**(1): p. 245-56.
127. Yakhnin, A.V., et al., *CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage.* Mol Microbiol, 2013. **87**(4): p. 851-66.
128. Doudoumis, V., 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): p. 4699.

129. Toh, H., 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): p. 149-56.
130. Clayton, A.L., et al., *A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses.* PLoS Genet, 2012. **8**(11): p. e1002990.
131. Kim, J.K., et al., *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**: p. 75-81.
132. Kim, J.K., et al., *Burkholderia gut symbionts enhance the innate immunity of host Riptortus pedestris.* Dev Comp Immunol, 2015. **53**(1): p. 265-9.
133. Attardo, G.M., et al., *Analysis of fat body transcriptome from the adult tsetse fly, Glossina morsitans morsitans.* Insect Mol Biol, 2006. **15**(4): p. 411-24.
134. Hu, C. and S. Aksoy, *Innate immune responses regulate trypanosome parasite infection of the tsetse fly Glossina morsitans morsitans.* Mol Microbiol, 2006. **60**(5): p. 1194-204.
135. González-Rete, B., et al., *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): p. 219.
136. Brown, M.J., Y. Moret, and P. Schmid-Hempel, *Activation of host constitutive immune defence by an intestinal trypanosome parasite of bumble bees.* Parasitology, 2003. **126**(Pt 3): p. 253-60.
137. Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a single report.* Bioinformatics, 2016. **32**(19): p. 3047-8.
138. Li, B., et al., *RNA-Seq gene expression estimation with read mapping uncertainty.* Bioinformatics, 2010. **26**(4): p. 493-500.
139. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
140. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
141. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
142. Tarazona, S., et al., *Differential expression in RNA-seq: a matter of depth.* Genome Res, 2011. **21**(12): p. 2213-23.
143. Huerta-Cepas, J., 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): p. D309-D314.
144. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

CHAPTER 4: Symbiont facilitated folate provisioning may function to regulate host gene expression in blood-feeding insects

Abstract

Epigenetics regulates the inheritance and expression of traits that are not due to changes in DNA sequences. DNA methylation is one of the most studied epigenetic mechanisms, and the B vitamin folate is a cofactor necessary for all biological methylation reactions. As bacteria are the providers of essential vitamins in several insect systems with restricted diets, particularly in blood-feeders, I hypothesize that epigenetics mediates interactions between the metabolic activities of insect hosts and their microbiota. However, knowledge of the mechanistic epigenetic links among host phenotype, metabolism and microbiome are scarce or missing. Here, I take a closer look at blood-feeding insects to review evidence on how epigenetics, specifically DNA methylation, influences the insect-microbiota interplay and, possibly their coevolution. First, I examine briefly the DNA methylation at its biochemical level, including the role of folate in DNA methylation. I comment on the context and genomic location of DNA methylation, and how this differs in insects relative to the more extensively studied vertebrate models. I describe the current knowledge regarding the involvement of microbiota in DNA methylation in insects. A special mention of the methylation in *Drosophila melanogaster* is included to highlight the controversy regarding methylation in Diptera. From here after, we will focus our attention on blood-feeding insects, given that their vitamin-restricted diet makes this group of organisms more dependent on symbiont provisioning, particularly regarding folate. We will address the predicted or empirical evidence regarding the microbiota as a possible source of folate. Also, we indicate the presence of a DNA methylation toolkit in the host genome and evidence regarding empirical detection of DNA methylation when available. We propose the hypothesis of a link between the folate provisioning roles of the microbiota and the healthy traits of blood-feeding insects, via the epigenetic mechanism of DNA methylation. We highlight key concepts to epigenetic in symbiosis and gaps in knowledge relevant to this interaction. Towards the end, we summarize common biological challenges relevant to the hypothesis and corresponding ways to address them. In the concluding section, we synthesize five main points in support of the views expressed.

Introduction

The symbionts of invertebrates with restricted diets

We refer to symbiosis in its original form, as a long-term physical association between different species, independent of the outcome [1,2]. Under this concept, the outcome of the association is a continuum that ranges from parasitism to mutualism. Bacteria are the most prevalent symbionts of Metazoa, namely, animals. We call these higher-order super-organisms holobionts, without implying any possible role as a target for natural selection [3]. Particularly, bacterial association with insects can be traced back to hundreds of millions of years [4]. During such long periods, insects and bacteria alike have co-diversified while exploiting countless niches in aquatic and terrestrial environments. Among several roles, microbes provide means for insects to subsist on otherwise harsh conditions. These partnerships have been termed acquisition events, where insects acquire the bacterial genomes [5] and henceforward the so called hologenome may work in a cooperative manner to make use of a particular resource. An ever-expanding research field points to the occurrence of independent acquisition events across insect lineages.

The most widely studied functional role of microbial symbiosis is nutrient provisioning of metabolite deficient diets. Insects that extract their food from xylem or phloem of plants [6] and the ones that feed on vertebrate blood are exemplary of this phenomenon. The acquisition of a symbiont is thought to be a continuous process, where a free-living bacterium forms an incidental physical association with the insect and its permanence is then selected for. The recently acquired symbionts resemble their free-living relatives from a genomics perspective. Originally, they can be found distributed in several tissues and their presence can be facultative, meaning that they are expendable to the host. However, when the association evolves to an inextricable one, the bacteria become localized to specialized host cells, called bacteriocytes, which in some orders form organs called mycetomes or bacteriomes; only the latter term will be used in this review, as it is the most common in recent scientific literature. The presence of these bacteria turns to be indispensable for the host, these are the so-called primary endosymbionts [7]. They commonly possess distinctive features such as vertical transmission and a drastically reduced size of the genome, characterized by a high coding density and adenine thymine bias [8]. On one hand, genes that the endosymbionts don't need for survival in the host environment degenerate and are eventually lost due to relaxed selection. On the other hand, genes that are pivotal to the relationship with their hosts are under purifying selection. The conservation of genes and pathways, such as those involved in vitamin synthesis, in an otherwise degenerate genome argues in favor of the importance of the corresponding enzymatic function being preserved [6].

Epigenetics in insects

Epigenetics controls gene expression and concomitant phenotype in differential ways that are not the consequence of changes in gene sequences [9]. Epigenetic mechanisms allow for a more rapid adaptation to changing environmental conditions, given that they fine tune the genomic potential to suit particular challenges [10]. Epigenetic changes are susceptible to being erased and reestablished in response to environmental cues[11–13]. Different phenotypes generated from a common genotype are referred to as polyphenisms [14]. Epigenetic mechanisms include microRNA, histones post-translational modifications, chromatin remodeling, and DNA methylation. These mechanisms are more thoroughly characterized in vertebrate models, but they are less understood within the invertebrates. Our focus will be on the latter, DNA methylation and particularly in a group of selected blood-feeding insects of epidemiological relevance. I use the term epigenetics in an inclusive manner, inheritable meaning passed on to daughter cells within an organism or to the offspring of an organism.

In insects, epigenetics will be expected to mediate responses to seasonal changes in food availability, changes in environmental stressors, such as temperature or humidity, and the presence of natural toxins in the environment[15–17]. This regulation would prove especially useful for relatively short-lived insects, which are expected to have several generations during a single year, with different seasons exposing successive generations to different environmental challenges. A time scale too short for DNA encoded protective mechanisms to evolve, but sufficient enough for epigenetics to intervene. Under this scenario, genes expected to be under epigenetic control would be associated with nutrient synthesis and storage, development, water content regulation, protein folding or detoxifying enzymes, for example.

The prevalence of genomic methylation differs across arthropod orders. Although, it is difficult cross-validate this pattern due to the lack of an exhaustive characterization of DNA methylomes

across insect taxa. Furthermore, methylation in arthropods seems to differ from the vertebrate counterpart from a localization and functional perspective. For example, in vertebrates, DNA methylation is more prevalent in promoter regions, and the modification is mainly associated with transcriptional regulation. Yet in insects, DNA methylation is more prevalent in introns and exons (gene bodies), where it is associated with gene splicing[18]. Also, DNA methylation in vertebrates is mostly associated with the CpG dinucleotide context, while in insects, such as Diptera, this modification is more prevalent in the CpA and CpT contexts[19].

DNA methylation is among the most amenable epigenetic modification to study given its susceptibility to automation. Commercial antibodies against methylated cytosine detect the presence of methylation in genomic DNA [20]. Bisulfite sequencing signals the specific position of a methylated base [21] and mapping the corresponding reads to a readily available insect genome allows to draw conclusions about the distribution of methylation in the genome. Furthermore, these studies permit the discovery of motifs that may be preferentially targeted for methylation [19,22]. Arguably, this will prove essential in upcoming insect DNA methylation research, given the evidence that suggests that insect whole-genome DNA methylation seems to have striking differences compared to the vertebrate ones. The generation of reference DNA methylomes for a variety of insect taxa in several developmental stages, and association of phenotypes with DNA methylation status will greatly facilitate the validation of epigenetic modifications.

The role of folate towards DNA methylation

Methylation requires folate for the generation of methyl groups to be transferred by methyltransferases. This vitamin is incorporated into the host cell and is sequestered by the addition of glutamyl groups (polyglutamate folate). Through successive steps, it carries the one-carbon group that will end up being used for cellular methylation reactions. First, the formate ion, mainly produced in the mitochondria, is attached to tetrahydrofolate to form formyl-tetrahydrofolate. This formyl group is converted to methenyl, and subsequently, to methyl. This methyl group is transferred to homocysteine to generate methionine. Methionine is converted to S-adenosyl-methionine (SAM), which is the substrate used by enzymes that catalyze the addition of methyl groups to the DNA [23,24]. Interestingly, levels of SAM are decreased in aposymbiotic tsetse flies bacteriomes [25].

Methylation of the DNA occurs at the 5’ position of the cytosine. Generally, in vertebrates it is found in the context of the CG dinucleotide, when it is called CpG, as a symmetrical modification. This dinucleotide tends to have an increased frequency in the promoter regions of genes, which are referred to as CpG islands. Here, methylation is thought to regulate the transcription of genes by means of altering the interaction with transcription factors and histones. A relatively hypermethylated status usually leads to a decreased transcription of the upstream gene, and vice versa, the hypomethylated status is associated with increased transcription. However, methylation patterns seem to differ greatly from these canonical locations and function [26] in different clades, particularly in insects [27]. For instance, within insects, the contexts CpA and CpT also show significant methylation. The body of genes, instead of promoter regions, seems to be a more prevalent target of methylation [28]. The percentage of global methylation in the genome is one or two orders of magnitude lower in insects, but, instead of silencing, the modification is more strongly associated with highly expressed genes [29,30]. Interestingly, the methylation in gene bodies of insects, such as bees, termites, and ants, has been shown to determine the expression of splice

variants, although the resulting transcriptomic and proteomic landscapes need to be further characterized [31–33].

A link between epigenetics and microbiota

Function of DNA methylation, RNA splicing

DNA methylation is related in different degrees to several functions in eukaryotes. In plants (*Oryza sativa*, *Arabidopsis thaliana*), algae (*Chlorella* sp., *Volvox carteri*), and vertebrates (*Tetraodon nigroviridis*), methylation in the region around the transcriptional start site is correlated with transcription. Highly methylated regions are associated with repression, while lower methylation levels correlate with increased expression. However, within invertebrates (the ascidian *Ciona intestinalis*, the starlet sea anemone *Nematostela vectensis*, the bee *Apis mellifera* and silk moth *Bombyx mori*) the correlation between methylation on transcriptional starting sites and transcription is lacking [34].

Particularly in insects, DNA methylation has a function in alternative RNA splicing. For instance, splice junctions are enriched for non-CpG methylation in bees [28] and different splice variants of the same gene are associated with the differential methylation of the gene [31]. Furthermore, differences in DNA methylation are hypothesized to be the cause behind reproductive caste determination in social insects, particularly, Hymenopteran such as bumblebees, ants, and bees. In bees, there is a methylation pattern exclusively associated with either the queen or the worker caste. The downregulation of the DNA methyltransferase 3 in the larval stage seems to determine the caste of the future adult, with the royal jelly functioning as the environmental cue driving cast determination [18].

Microbiota influences host methylome

The bacterium *Wolbachia* is a sexual parasite widely prevalent in arthropods. It is transmitted vertically by the mother and it induces bias in the sex ratio of the offspring towards a higher proportion of females [35]. The bacterium possesses intricate ways to manipulate host methylation. In mosquitoes, for example, *Wolbachia* induces the production of a microRNA that regulates the expression of the only host-encoded DNA methyltransferase, *dnmt2* [36].

It has been proposed that microbiota is able to alter the methylation landscape in symbiotic systems such as plants [37], lichens [38], anemones [39], mice [40,41] and insect hosts, with concomitant changes in phenotype. For instance, the sexual parasite *Wolbachia* induces feminization in the leafhopper *Zyginidia pullula* by disruption of the male imprinting. *Wolbachia* is vertically transmitted through the germline of females, consequently an increased rate of female offspring favors the spread of this bacteria in the population. During feminization, the offspring that is male, according to their genetics, is transformed into females. The transformed offspring, called intersexes, vary according to the degree in which they are affected. Some offspring only resemble the female external morphology, while keeping the testes. However, some male individuals exhibit ovaries. The methylation imprinting pattern of feminized males with testes resembles the pattern observed in regular males. Interestingly, in feminized males with ovaries, the male methylation imprinting pattern is lost, so the methylation landscape becomes indistinguishable from that of females [42].

These examples imply that there must be a strong pressure to use epigenetics, DNA methylation in particular, to coordinate the interaction between host and symbiont. This becomes evident from (1) its usage during infections by the ancient parasite *Wolbachia*, (2) its utilization in different insect orders, and (3) the evidence that insects may deploy DNA methylation towards microbial interactions.

Study of symbiont provisioning role

There are three approaches to study the provisioning role of symbionts in insects [44]. First, in dietary supplementation experiments, the aposymbiotic hosts are reared with minimum media of a known composition. When phenotypic abnormalities are observed, specific nutrients of interest are added back to the media to look for restoration of the regular physiological function, for instance, fecundity, development, or immunity. Second, the analysis of biomarkers determines the status or concentration of key metabolites that belong to a particular pathway of interest, such as the vitamins themselves, or intermediates in a pathway, such as homocysteine or SAM in the case of DNA methylation. Third, the genomic approach search for the predicted capabilities of host and symbionts in an effort to find complete, partial, or missing metabolic pathways and see how or if they can interact in such a way that symbionts may provide the nutrient of interest. The techniques mentioned previously find limitations, depending on the system being studied. For instance, the host may have a specialized diet that cannot be easily replaced by minimum media, or the apo-symbiotic form of the host may not be viable for long. The metabolites may lack straightforward methods to be detected and lastly, the genomic tactics rely on fully sequenced and annotated genomes from host and symbionts. Consequently, the most solid approaches incorporate the different methods to better asses the actual biological capabilities of the partners.

The genomic approach in the study of the provisioning role of bacterial symbiont in insects

The genomic approach has been used to analyze the great available number of symbiont genomes looking for signs of vitamin synthesis and provisioning [75], including endosymbionts, which live within the body cavity of arthropods, and closely related free-living bacteria. Interestingly, Alpha, and Gammaproteobacteria, and Tenericutes represent taxa with a significantly greater proportion of members having at least a partial pathway for the synthesis of folate. This is particularly relevant, given that these taxa include symbionts specifically found in blood-feeding arthropods, such as the black legged tick *Ixodes scapularis*, the lone star tick *Amblyomma americanum*, the body louse *Pediculus humanus corporis*, the mosquito *Culex pipiens*, the bed bug *Cimex lectularius* and the tsetse fly *Glossina morsitans*. If a closely related free-living bacteria has the synthetic capability for a given vitamin, an endosymbiont will be more likely to conserve in its genome at least the corresponding partial pathway. Strikingly, B vitamins, B1-B9, account for most of the pathways where vitamins are made available to the host metabolic processes. They also speculate that partial pathways enable symbionts to use intermediate metabolites they salvage from the environment to produce the finished vitamin. Although the authors found no association between host taxa and the capability of a symbiont for vitamin synthesis, it would be interesting to take a similar approach that would incorporate host biology, specifically feeding habits, such as composition or source of their diet, because diet, along with host metabolism, maybe more important determinants of the intermediate metabolites pool available to the symbiont.

Support for DNA methylation in Diptera

The existence of a functional DNA methylation system, and DNA methylation altogether, in *D. melanogaster* has been intensely questioned. However, there is incontrovertible evidence from different researchers and experimental approaches that support DNA methylation in the fly [45,46]. Although the identity of the enzyme remains debated, it is known that (1) There is 5mC [47]. The presence of 5mC in the DNA has been detected by high performance liquid chromatography and thin layer chromatography [48]. Success in detection of DNA methylation is attributed to an improvement in sensitivity of the methods applied, given the extremely low percentage of methylated cytosine in flies. (2) The DNA is methylated. Specific antibody detection has been achieved in slot blots with appropriate controls, i.e., the cow *Bos taurus* as positive and the yeast *Saccharomyces cerevisiae* as negative control[49,50]. Also, antibody-mediated enrichment of methylated DNA fractions has been performed as a selectivity step prior to high throughput sequencing [19]. (3) A functional DNA methylation machinery is present in flies. The identity of the physiologically active DNA methyltransferase in the fly remains unclear. RNA interference and inhibition assays implying DNA methyltransferase 2 in DNA methylation [50]. However, enzymatic *in vitro* studies that point to RNA methylation function, cytoplasmic localization, and structural constrains that would halt its interaction with DNA argue against a physiologically relevant role for DNA methyltransferase 2 in DNA methylation [51]. The existence of an uncharacterized enzyme has been also suggested [19]. However, the DNA methyltransferase activity itself is evident. The presence of the modified base is the first indication. Furthermore, the enzymatic activity was confirmed using microarray. A microarray containing 14 K oligos was designed and tested with a *D. melanogaster* cellular extract. The researchers verified the addition of methyl groups to specific oligos, in the presence of SAM, methyl donor for methylation reactions [22]. (4) New characteristic DNA methylation motifs have emerged. The position of the modified cytosines has been mapped genome wide in such a detail, that it has been possible to deduce 4-7 base long DNA motifs where methylation is likely to occur. Surprisingly, this methylation is more prevalent in the CpT and CpA contexts, and it was also found in CpC context. The latter was verified by *MspI/HpaII* mediated PCR. *MspI* and *HpaII* are isoschizomers, meaning that they are enzymes that recognize the same target sequence, in this case, 5'-CCGG-3'. Methylation of the internal cytosine inhibits cleavage by *HpaII*, but not by *MspI* [52]. Methylation at the external cytosine in the sequence 5'-CCGG-3' inhibits cleavage by both enzymes [22]. If the sample DNA is incubated with both enzymes simultaneously and there is no cleavage, this indicates the presence of methylation in the external cytosine. Consequently, if primers anneal at a region containing this sequence, the gene will be amplified only if the region is intact, indicating methylation [19,22]. This exhibits a striking contrast with the predominant CpG methylation observed in vertebrates.

***Glossina morsitans* (Diptera: Glossinidae)**

The tsetse flies *Glossina* spp. are the vectors of African trypanosomiasis [53], a debilitating disease that, in its human and animal forms, has become a severe health and economic problem in Sub-Saharan Africa. This fly has an exclusive hematophagous habit, and it feeds on endemic vertebrate fauna, cattle, and humans. Blood is particularly deficient in the B Vitamin folate (Greiner et al., 1978; Pietrzik et al., 2010). The tsetse microbiota is relatively simple. It is constituted by two endosymbiotic gamma-proteobacteria, the obligate primary endosymbiont *Wigglesworthia glossinidia*, localized within the bacteriome [57–59] and the facultative *Sodalis glossinidius* [60]. Two parasites found in tsetse also include *Trypanosoma* and the sexual parasite *Wolbachia* [61]. The elucidation of the provisioning role of *Wigglesworthia* in the tsetse is exemplary of the use of three

approaches mentioned earlier, namely, genomic, dietary, and metabolite status. Genome sequences are available for both tsetse [62] and *Wigglesworthia* [63,64]. Bioinformatics indicates that from a genomic perspective, *Wigglesworthia morsitans*, endosymbiont of *Glossina morsitans,* is capable of synthesizing folate [64]. This is based on the conservation of functional genes that map and permit the reconstruction of a complete folate synthesis pathway. The corresponding enzymes encoded by these genes mediate the synthesis of chorismate from phosphoenolpyruvate and erythrose 4-phosphate. Chorismate may be used by the bacterium to produce p-aminobenzoate, which is then incorporated in the synthesis of folate [65]. The prediction is supported by the increased transcriptional activity of several genes in this pathway, especially during pregnancy. Dietary supplementation studies backed up the contribution of *Wigglesworthia* [66]. When the bacterium is knocked out of the partnership, the fly health shows negative effects, such as decreased immunological response [67], lower fecundity, and delayed larval development. Supplementing blood meals with folate rescues regular phenotypes in these aposymbiotic flies, such as fecundity, larval weight, and development. Accordingly, folate status measurements show that the vitamin is higher in bacteriomes, where *Wigglesworthia* is located, and importantly it is higher during pregnancy, which, according to fly reproduction biology, includes gametogenesis and larval development. Also, concomitant with the loss of the symbiont, there is a significant decrease in SAM, the universal substrate for methylation [68].

G. morsitans is predicted to have only DNA methyltransferase 2 (GMOY008571) with orthologs across other *Glossina* species, namely, *G. austeni* (GAUT008339), *G. brevipalpis* (GBRI003588), *G. pallidipes* (GPAI047253), *G. fuscipes* (GFUI001419) and *G. palpalis* (GPPI011588). All of these species are also constituted of female and male exclusive blood feeders.

To our knowledge, there are no studies searching for empirical evidence on methylation in the tsetse genome. Due to the lack of DNA methyltransferases 1 and 3 in the genome, this fly has been predicted to lack DNA methylation [69]. However, due to its close evolutionary relation to *D. melanogaster* and the extensive characterization of DNA methylation in the fruit fly [19,22], we are inclined to hypothesize the presence of methylation in the genomic DNA of the tsetse, albeit likely at low percentages. Expression of putative DNA methytransferase orthologs into insect cells otherwise devoid of DNA methylation activity may prove a suitable way to validate the intrinsic capacities of the enzymes. Furthermore, preliminary results in our lab show strong indications towards the presence of methylated DNA in tsetse. Immunoblot against methylated cytosine shows different degrees of methylation across total DNA from tsetse in several developmental stages (Fig. 24). The assay included *Bos taurus* DNA as a positive control for DNA methylation and *Saccharomyces cerevisiae* as a negative control. An equal amount of DNA of the indicated samples was loaded in each slot, and the membrane was incubated with antibody directed toward 5-methyl-cytosine (α5mC). Development revealed bands of varying intensities depending on life stage. More assays are needed to distinguish tsetse DNA methylation to possible interference from methylation on microbial DNA. A suggested approach to study the involvement of symbionts in determining the DNA methylation landscape of blood-feeding symbionts via folate provisioning is presented in Figure 25 and summary of the DNA methylation status and microbiome of main blood-feeding arthropods affecting humans is presented in Table 6.

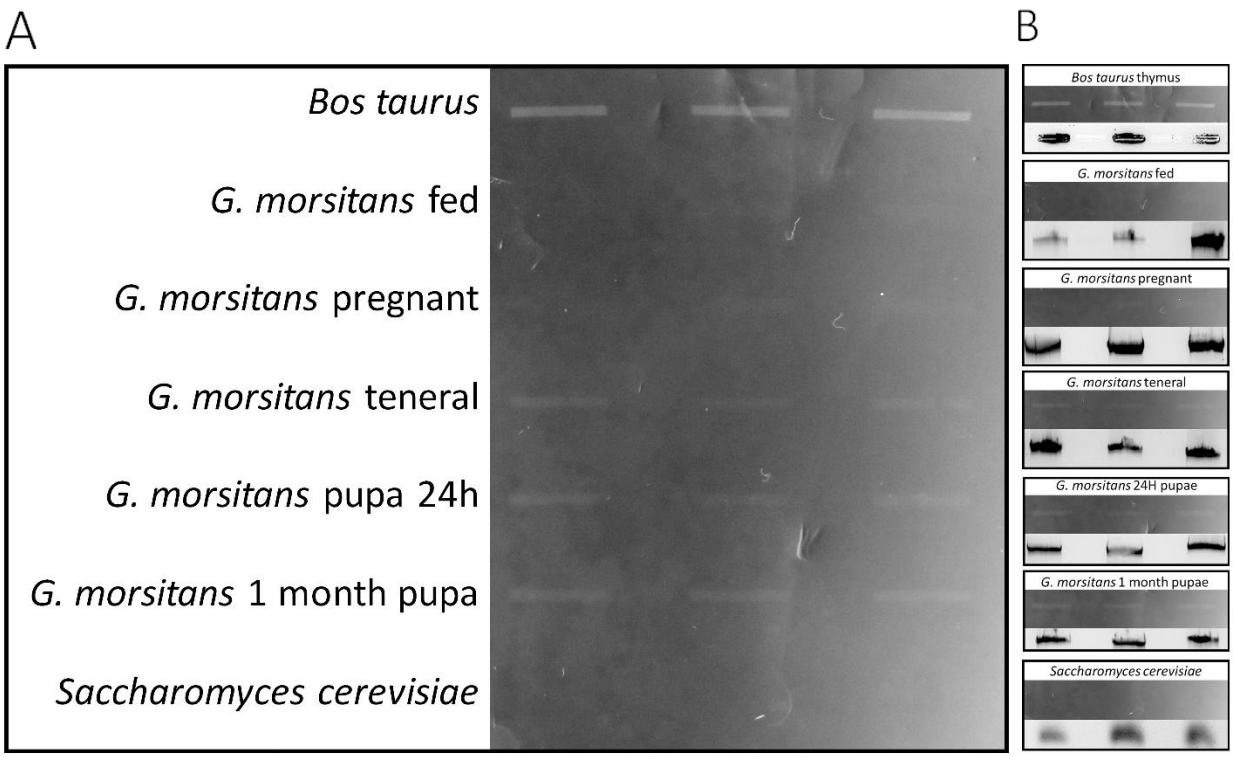

Figure 24. Evidence of DNA methylation in total DNA from tsetse flies.
A. Immunoblot where 3500ng DNA of indicated sample were loaded in each slot. B. On each panel: Top: Slot blot stained via α5mC antibody; bottom: DNA loading control where the same amount of DNA was stained with Ethidium Bromide on a 1.5% agarose gel.

***Rhodnius prolixus* (Hemiptera: Reduvidae)**

The kissing bug *Rhodnius prolixus* is an obligate blood-feeder. This insect is the vector of *Trypanosoma cruzi*, the etiologic agent of Chagas disease, which causes heart failure [70]. Symbionts were implicated in folate provisioning role in early studies demonstrating that they produce an excess of folate that could be relayed to the host [71]. The insect has a diverse microbiota, including Gammaproteobactria and Actinobacteria [72]. Within the latter group, the insect harbors in its hindgut the extracellular bacterium *Rhodococcus rhodnii*, which is transmitted to the offspring via coprophagy [73]. This symbiont has a hypothesized participation in the B vitamin metabolism of the insect [73] and the predicted capability of folate synthesis [65].

DNA methyltransferase 2 (RPRC014336-PA) is present in the *R. prolixus* genome [74]. Comparative *in silico* studies between the genomes of the kissing bug and the closely related aphid *Acyrthosiphon pisum* argue in favor of a functional methylation system. *A. pisum* has methylation in its genome [75] and possesses both DNA methyltransferases 1 and 3, while *R. prolixus* have DNA methyltransferase 1 (RPRC000633-PA), but not 3 [76]. The kissing bug has predicted CpG methylation [69], corroborated by empirical detection of methylated DNA[77].

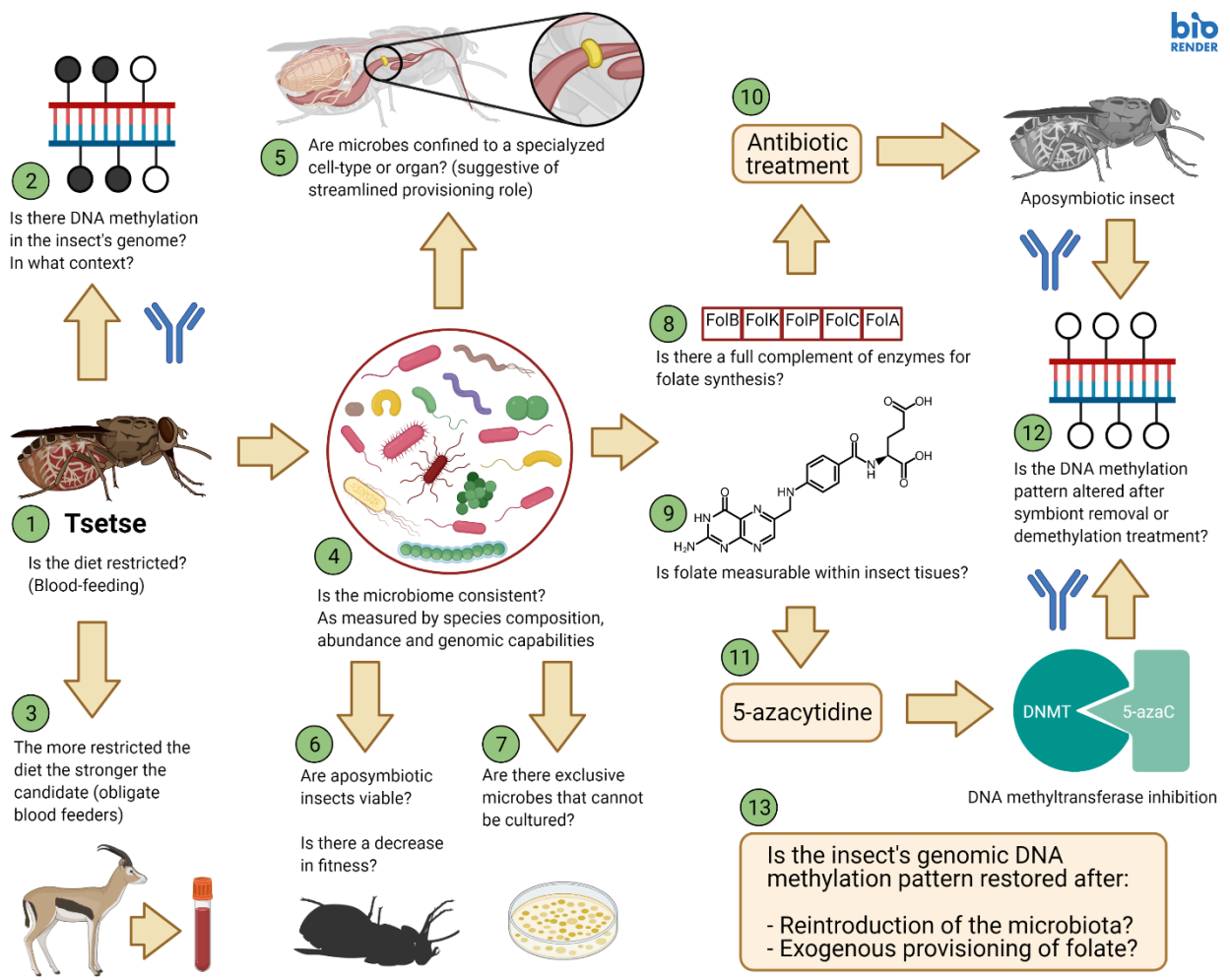

Figure 25. Suggested experimental flow to study the impact of the microbiome's folate provisioning role in host genome methylation.

① This approach would focus on insects with a restricted diet, more specifically hematophagous insects whose food source is poor in folate. ② The impact of folate provisioning on genome methylation has to be preceded for the selection of an appropriate technique to detect 5mC and determine the context (i.e., CpG or CpA) and genomic (i.e., intron, exon, non-coding region) preferences. The antibody icon next to the arrow indicates that an enrichment step targeting the methylated fraction of the genome should be used to improve detection. This step may be mediated by α-5mC antibodies. ③ The ecological context of the insect vector should be considered, particularly regarding its natural food sources. This step is indicated in order to support the hypothesis that the blood folate content of the natural prey would be too low to support insect physiology. ④ Consistency of the members of the microbiome and their genomic capabilities across different populations of the vector is a strong indicator of the co-evolution that occurs in insect-bacterial symbiosis, and that leads to microbiota fulfilling provisioning roles. ⑤ The segregation of specific bacterial populations to a dedicated cell type or organ points to selective pressures to keep the bacterium within an ideal microenvironment. This positions the bacterium as an ideal candidate for essential physiological roles. ⑥ The vital role of candidate microbes for the insect can be further tested by forcefully disrupting the symbiotic association. A decreased fitness in the aposymbiotic population demonstrates that the bacterium is vital for species survival. ⑦ An equivalent approach would demonstrate a long history of coevolution from the side of the bacterium: If the candidate member of the microbiota is incapable of survival outside

its insect host, this will indicate that its genome and/or physiology is streamlined for very specific functions within the insect. ⑧ The finding of folate synthesis capabilities within the genome of the candidate bacterium is strongly supportive of a provisioning role. ⑨ Furthermore, detection of folate, especially if the vitamin's levels are tied to bacterial presence, location or metabolic state; or to the physiological state of the host; is a promising indicator of a tightly regulated metabolic exchange. ⑩ To study the impact of the microbial candidate on the host genome a course similar to the study of the folate provisioning role can be followed: the disruption of the symbiosis via antibiotic treatment, or other means, would have a measurable effect on the 5mC content and location if the symbiont is involved in host genome regulation via DNA methylation. ⑪ In concordance with the previous step, a demethylation treatment would also lead to a disruption of the DNA methylation landscape. ⑫ The outcome of each one of the two previous steps would be a DNA methylation pattern that is significantly different from the landscape observed in the first step in this experimental flow. ⑬ Confirmation of the direct involvement of the microbiome and DNA methyltransferase activity in determining the insect host DNA methylation landscape will take the form of rescue experiments: the microbiome may be reconstituted in aposymbiotic insects or, similarly, folate may be provided via exogenous supplementation. The expected result would be the restoration of a DNA methylation landscape similar to that present in natural insect populations.

***Cimex lectularius* (Hemiptera: Cimicidae)**

The bed bugs *Cimex lectularius* are an increasingly common insect pest. They are not known to be vectors for human parasites, although it is associated with dermal and psychological discomfort, and also to economic loss due to decreased reputation and furniture replacement in the hospitality industry [78]. Both males and females are obligate blood-feeders. This insect feeds on humans, although it has been found associated with bats and birds in the wild [78]. Bed bugs have bacteriomes that harbor *Wolbachia*. *Wolbachia* in the bed bug localizes primarily to the bacteriomes and to the ovaries, notoriously in the nurse cells that connect to the developing oocyte. Within the oocyte *Wolbachia* abounds in the region that will give rise to the germline, evidence supporting vertical transmission. Removing the symbiont has especially detrimental consequences during early developmental stages; for instance, eggs exhibit a darkened color and abnormal shape. Adult emergence is also negatively affected. A dietary supplementation approach to aposymbiotic bed bugs shows that *Wolbachia* is the main source of B Vitamins for this insect [79]. A Gammaproteobacteria yet to be identified is also found in bacteriocytes and other tissues, such as Malpighian tubes. These symbionts are vertically transmitted by the mother. *Wolbachia*, unlike its predominantly parasitic role in other insects, has a provisioning role for *C. lectularius*. *Wolbachia* provides the B Vitamins Biotin (B7) and Riboflavin (B2). The biosynthetic pathway for folate is partially retained in this *Wolbachia*, leading to the hypothesis that the bacterium may take part in the folate metabolism in this system[65,80]. A folate provisioning role for the unidentified symbiont could still be speculated.

C. lectularius has a predicted DNA methyltransferase 1 (CLEC007142-PA). DNA methyltransferase 3 gene seems to be absent, but DNA methyltransferase 2 is present (CLEC007374-PA). The bed bug has predicted CpG methylation [69], to our knowledge, there are no studies searching for empirical evidence on methylation in the bed bug genome.

***Ctenocephalides felis* (Siphonaptera: Pulicidae)**

Fleas are insects in the order Siphonaptera, with approximately 2,500 species [81]. They bare a high medical relevance as vectors of etiologic agents such as *Yersinia pestis*, which causes the plague, and *Rickettsia felis*, which causes cat-flea typhus in humans. Although exclusive blood feeders as

adults, larval stages of the flea feed on adult feces and other detritus found in their host environment [82]. However, little seem to have been researched on vitamins requirements or sources for fleas, specifically folate. Fleas have a highly diverse microbiota [83–85]. Interestingly, flea microbiota in these studies, coming from USA, Uganda, and Australia, respectively, are dominated by Alphaproteobacteria, Gammaproteobacteria, and Tenericutes, which are the bacterial taxa with the greatest number of endosymbionts possessing at least a partial pathway for folate synthesis [44]. Particularly, *Wolbachia* is among the most prevalent symbionts. Although generally a sexual parasite, it is worth noting the B Vitamin provisioning role it serves in another hematophagous insect, the bed bug *C. lectularius* [79].

Preliminary data suggest the presence of DNA methyltransferases 1 (XP_026477759.1 and XP_026471107.1) and 2 (XP_026480230.1) in the cat flea genome[86]. Studies are being undertaken to further characterize these genes. To our knowledge, there are no studies searching for empirical evidence on methylation in the flea genome.

Pediculus humanus corporis

There are approximately 550 species of lice (Phthiraptera: Anoplura). They are obligate blood-feeding ectoparasites on mammals. The body louse *Pediculus humanus corporis* is medically relevant, as it feeds only on humans and constitutes the vector for the infectious diseases louse-borne relapsing fever, trench fever and epidemic typhus [87,88]. The body louse harbors the endosymbiotic Gammaproteobacterium *Riesia pediculicola* in bacteriomes [89]. This symbiont is also vertically transmitted by the females. Interestingly, *R. pediculicola* retains in its genome synthetic capabilities for folate [87,90,91].

The human body louse genome [90] has also a predicted DNA methyltransferase 1 (XP_002431878.1 and XP_002432160.1) and predicted CpG methylation [69]. DNA methyltransferase 2 is present (XP_002432555.1). To our knowledge, there are no studies searching for empirical evidence on methylation in the lice genome.

Culicidae

The mosquito *Aedes aegypti* (Diptera: Culicidae) is the vector of viral, bacterial, and protozoan etiological agents. The larval stages and pupa are aquatic. Larvae feed on algae and other microorganisms found in their habitat. The adults feed on nectar, but only the females feed on blood prior to oviposition [92]. *Wolbachia* infection alters the methylation landscape of mosquito genome leading to an overall hypomethylated state, where the demethylation of 699 genes is found, compared to non-infected controls, while only methylation of 63 genes is induced [93]. *Wolbachia* also induces the production of a microRNA that targets the DNA methyltransferase 2 (XM_001657505) transcript, the only DNA methyltransferase found in the genome [94]. A congruent model was proposed where *Wolbachia* induces the host microRNA production, which in turn downregulates DNA methyltransferase 2 and causes a hypomethylated genome[36]. Interestingly, suppression of DNA methyltransferase 2 is necessary for *Wolbachia* proliferation and seems to be detrimental for the dengue virus. Conversely, overexpression of DNA methyltransferase 2 is deleterious for *Wolbachia*, but enhances viral replication. *Aedes aegypti* also has a diverse gut microbiota that changes significantly when insects transition to from sugar feeding to blood feeding [95].

Mosquitoes are predicted to have only DNA methyltransferase 2 [69]. Other genera possess homologs such as that of *Anopheles gambiae* (AGAP004101) [45,96]. There is also empirical evidence for methylation in *Aedes albopictus* [97,98]. These studies detected the presence of methylated cytosine in the DNA [98] and localized the methylation preferentially to the centromeres in *A. albopictus* [97]. Also, methylation status in the region surrounding promoters was investigated in *A. aegypti* [93], however, the prevalent context for methylation, namely CpG, CpA or CpT, is not provided.

This system establishes a strong argument linking methylation of a blood feeding insect host genome and the survival of a symbiotic bacteria. Given the diversity of insects and their bacterial associations, it is plausible to hypothesize that bacteria could develop diverse ways to exert control on host genome methylation, besides microRNA, such as regulating folate provision.

Acari

Ticks are arthropods of the Subclass Acari in the Class Arachnida. They are obligate ectoparasites of blood-feeding habit. Ticks have a high vector potential and are responsible for the transmission of the agents of infectious diseases such as Babesiosis and Lyme disease. The host range of the approximately 900 species of ticks encompass mammals, reptiles, birds, and amphibians [99,100]. Ticks possess a diverse microbiota [101,102]. Particularly interesting is the *Coxiella*-like endosymbiont found in *Amblyomma americanum*. This endosymbiont is not located in bacteriocytes as other endosymbionts in blood-feeding insects, which may be suggestive of a recently acquired symbiont transitioning between lifestyles, as proposed for *Sodalis glossinidius* within tsetse [103]. Removing the *Coxiella*-like endosymbiont has detrimental effects towards the reproduction of *A. americanum* [104], thus, the genomic approach has been used to imply a provisioning role for this bacterium. First, the bacterium lacks virulence genes found in other *Coxiella* and related species. Second and most important, the complete pathway for folate synthesis is present [105]. In ticks, both *Coxiella* and *Rickettsia* symbionts are vertically transmitted by mothers [65]. However, neither *Rickettsia buchneri*, found in the black legged tick *Ixodes scapularis*, nor *Rickettsia* species found in the Western black legged tick *Ixodes pacificus* [106], have *folB*, which is required for tetrahydrofolate biosynthesis. Consequently, the role of *Rickettsia* species in folate biosynthesis and provisioning has been rejected [107]. The genome of the closely related tick *Ixodes scapularis* [108] has a predicted DNA methyltransferase 1 (XP_002403216.1) and a DNA methyltransferase 2 (ISCW023659-RA).

There is evidence for the presence of methylation in the tick genome. Digestion with the methylation sensitive enzyme *HpaII* yields a pattern that is different from the undigested genomic DNA. Furthermore, probes targeting methylated DNA localize methylated regions to heterochromatic pericentromeric parts of the chromosomes [109]. However, to our knowledge, the influence of the microbiota on tick genomic DNA methylation has not been studied.

Table 6. Summary of the DNA methylation status and microbiome of main blood-feeding arthropods affecting humans.

Arthropod	Blood source	Symbiont with folate provisioning potential	Genomic DNA methylation	Predicted DNMTs	Most common pathogens vectored
***Glossina morsitans* (Diptera: Glossinidae)**	Vertebrates	*Wigglesworthia*	Unknown	DNMT2	Old World Trypanosomes
***Rhodnius prolixus* (Hemiptera: Reduvidae)**	Vertebrates	*Rhodococcus rhodnii*	Yes	DNMT1 DNMT2	New World Trypanosomes
***Cimex lectularius* (Hemiptera: Cimicidae)**	Vertebrates	*Wolbachia*	Unknown	DNMT1 DNMT2	None known
***Ctenocephalides felis* (Siphonaptera: Pulicidae)**	Mammals	Diverse microbiota dominated by Alphaproteobacteria, Gammaproteobacteria and Tenericutes	Unknown	DNMT1 DNMT2	*Yersinia pestis*, *Rickettsia felis*
Pediculus humanus corporis	Humans	*Riesia pediculicola*	Unknown	DNMT1 DNMT2	*Borrelia recurrentis*, *Bartonella quintana*, *Rickettsia prowazekii*
***Aedes spp* and *Anopheles spp* (Diptera: Culicidae)**	Mammals	Diverse microbiota depending on developmental stage and feeding status	Yes	DNMT2	Viral, bacterial, and protozoan etiological agents
Acari (Class Arachnida)	Vertebrates	None known	Yes	DNMT1 DNMT2	*Borrelia burgdorferi*, *Babesia microti*

Challenges and strategies

We acknowledge that there are several milestones to be overcome in order to imply the folate as a key mediator in the collaboration between bacterial symbionts and insect hosts, particularly regarding gene regulation via DNA methylation in blood feeders. However, available resources offer ways to tackle the gaps in knowledge. To briefly list the issues, let us consider the following. The association of a core microbiome with the particular insect of interest has to be addressed. Metagenomic approaches permit the identification of the most prevalent members of the microbiome. The intimacy of the host-microbe relationship is one of the most relevant factors. This can be addressed with observations on the location and transmission of the symbionts; where localization to bacteriome and vertical transmission are strongly suggestive clues, respectively. The potential for folate synthesis in the symbiont must be studied. This is accomplished by comparative and predictive analyses based on the sequenced genome of the symbiont. A complementary approach includes measurement of key metabolite status, such as the concentration of folate or SAM. Another factor, beyond the scope of this review, is the capability of the symbionts to generate an excess of folate and relay it to the host. To this end, predicted transporters in symbiont genomes can be compared to bacterial transporters of known capabilities to predict which, if any, is involved in folate transport.

From the host perspective, equivalent challenges should be addressed. Namely, the presence of DNA methylation. Available insect genomes are to be analyzed in the search for a DNA methylation toolkit. Possession of a set of the three canonical DNA methyltransferases is the strongest clue to methylation. However, as we have seen, the lack of DNA methyltransferases 1 and 3 does not necessarily exclude the presence of genomic methylation, as DNA

methyltransferase 2 or a hypothetical not characterized enzyme may perform this function. Complementary, the detection of methylated cytosines in the DNA proves indispensable as a starting point for the elucidation of the effect of this modification in gene expression and phenotypic alterations of the host. Bisulfite sequencing and immunological detection are commonly used means to this end. We encourage caution when using software to predict the presence of methylation in the host genome. Studies based on such approaches disregard the presence of methylation in epidemiologically relevant vectors and model system such as the Dipteran *Aedes aegypti*, *Aedes albopictus*, *Anopheles gambiae* and *Drosophila melanogaster*, despite decades of accumulated evidence supporting methylation in their genomic DNA [69,110].

Conclusion

Synthesizing, five main clues point to a possible intervention of folate production by symbionts into the epigenetics of blood-feeding hosts. First, it is costly for a symbiont to keep metabolic pathways. B vitamins synthetic pathways, such as folate, are retained as partial or complete, but functional, in several blood-feeding host symbionts. This is especially noteworthy, given the tendency in endosymbionts to decrease their genome size and degenerate genes due to relaxed selection. Secondly, it seems to be costly for the host to keep symbionts. For instance, under favorable laboratory conditions, as seen in tsetse flies, a host deprived from its symbiont may have an increased lifespan, pointing to a reallocation of nutrients that is not needed when the symbiont is not present. In this specific case, the symbiont is known to be the major source of folate [66]. Third, B vitamins, folate, in particular, is needed by insects, such as fruit flies [111] and tsetse [66], in order to have normal fecundity and larval development. Fourth, epigenetic mechanisms, DNA methylation, in particular, have their peak activity during gametogenesis and early development, when methylation patterns are being erased and reestablished, imprinting is being protected, and DNA methyltransferases are transcribed most abundantly. Fifth, folate is essential for the production of SAM, the substrate for DNA methylation reactions, and the sole food source of the blood-feeding hosts is notorious for lacking this particular vitamin. Taken together these points of evidence offer an arguably plausible flow of events where the symbiont provides folate, which is transported and incorporated into the host metabolism. The vitamin would be used for the establishment of a DNA methylation landscape, which, in turn, will ensure a correct gene expression. The overall consequence would be the healthy fecundity and early development observed in host populations. This may also prove advantageous to the symbiont, as a healthy host population provides the necessary biotic environment and means for the symbiont to spread in a sustained manner.

References

1. Frank B. Über die auf Wurzelsymbiose beruhende Ernährung gewisser Bäume durch unterirdische Pilze from Berichte der Deutschen Botanischen Gesellschaft. *Mycorrhiza*. 1885;3:128-145. doi:10.1007/s00572-004-0329-y
2. de Bary A. *Die Erscheinung Der Symbiose. Vortrag Gehalten Auf Der Versammlung Deutscher Naturforscher Und Aerzte Zu Cassel*. Verlag von Karl J. Trubner; 1879.
3. Moran NA, Sloan DB. The Hologenome Concept: Helpful or Hollow? *PLoS Biol*. 2015;13(12):1-10. doi:10.1371/journal.pbio.1002311
4. Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: Insights from genomics. *Trends Ecol Evol*. 2000;15(8):321-326. doi:10.1016/S0169-5347(00)01902-9
5. Margulis L, Sagan D. *Acquiring Genomes: A Theory Of The Origin Of Species*. 1st ed. Basic Books; 2003.
6. Moran NA, Plague GR, Sandström JP, Wilcox JL. A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proc Natl Acad Sci U S A*. 2003;100 Suppl:14543-14548. doi:10.1073/pnas.2135345100
7. Dale C, Moran NA. Molecular Interactions between Bacterial Symbionts and Their Hosts. *Cell*. 2006;126(3):453-465. doi:10.1016/j.cell.2006.07.014
8. Moran NA, Baumann P. Bacterial endosymbionts in animals. *Curr Opin Microbiol*. 2000;3(3):270-275. doi:10.1016/S1369-5274(00)00088-6
9. Choudhuri S. From Waddington's epigenetic landscape to small noncoding RNA: Some important milestones in the history of epigenetics research. *Toxicol Mech Methods*. 2011;21(4):252-274. doi:10.3109/15376516.2011.559695
10. Vogt G. Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: insights from clonal, invasive, polyploid, and domesticated animals. *Environ Epigenetics*. 2017;3(1):1-17. doi:10.1093/eep/dvx002
11. Lucas K, Raikhel AS. Insect MicroRNAs: Biogenesis, expression profiling and biological functions. *Insect Biochem Mol Biol*. 2013;43(1):24-38. doi:10.1016/j.ibmb.2012.10.009
12. Deans C, Maggert KA. What do you mean, "Epigenetic"? *Genetics*. 2015;199(4):887-896. doi:10.1534/genetics.114.173492
13. Tammen SA, Friso S, Choi SW. Epigenetics: The link between nature and nurture. *Mol Aspects Med*. 2013;34(4):753-764. doi:10.1016/j.mam.2012.07.018
14. Simpson SJ, Sword GA, Lo N. Polyphenism in insects. *Curr Biol*. 2011;21(18):R738-R749. doi:10.1016/j.cub.2011.06.006
15. Bind MA, Zanobetti A, Gasparrini A, et al. Effects of temperature and relative humidity on DNA methylation. *Epidemiology*. 2014;25(4):561-569. doi:10.1097/EDE.0000000000000120
16. McCaw BA, Stevenson TJ, Lancaster LT. Epigenetic Responses to Temperature and Climate. *Integr Comp Biol*. Published online 2020:1-12. doi:10.1093/icb/icaa049
17. Villagra C, Frías-Lasserre D. Epigenetic Molecular Mechanisms in Insects. *Neotrop Entomol*. 2020;49(5):615-642. doi:10.1007/s13744-020-00777-8
18. Li-Byarlay H. The Function of DNA Methylation Marks in Social Insects. *Front Ecol Evol*. 2016;4(May):1-8. doi:10.3389/fevo.2016.00057

19. Takayama S, Dhahbi J, Roberts A, et al. Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. *Genome Res*. 2014;24(5):821-830. doi:10.1101/gr.162412.113
20. Kunert N. A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development*. 2003;130(21):5083-5090. doi:10.1242/dev.00716
21. Ku CS, Naidoo N, Wu M, Soong R. Studying the epigenome using next generation sequencing. *J Med Genet*. 2011;48(11):721-730. doi:10.1136/jmedgenet-2011-100242
22. Panikar CS, Rajpathak SN, Abhyankar V, Deshmukh S, Deobagkar DD. Presence of DNA methyltransferase activity and CpC methylation in Drosophila melanogaster. *Mol Biol Rep*. 2015;42(12):1615-1621. doi:10.1007/s11033-015-3931-5
23. Crider KS, Yang TP, Berry RJ, Bailey LB. Folate and DNA Methylation : A Review of Molecular Mechanisms and the Evidence for Folate ' s Role. *Am Soc Nutr*. 2012;3(14):21-38. doi:10.3945/an.111.000992.Figure
24. Shorter KR, Felder MR, Vrana PB. Consequences of dietary methyl donor supplements: Is more always better? *Prog Biophys Mol Biol*. 2015;118(1-2):14-20. doi:10.1016/j.pbiomolbio.2015.03.007
25. Bing X, Attardo GM, Vigneron A, et al. Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proceedings Biol Sci*. 2017;284(1857):20170360. doi:10.1098/rspb.2017.0360
26. Suzuki MM, Bird A. DNA methylation landscapes: Provocative insights from epigenomics. *Nat Rev Genet*. 2008;9(6):465-476. doi:10.1038/nrg2341
27. Head JA. Patterns of DNA methylation in animals: An ecotoxicological perspective. *Integr Comp Biol*. 2014;54(1):77-86. doi:10.1093/icb/icu025
28. Cingolani P, Cao X, Khetani RS, et al. Intronic Non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees. *BMC Genomics*. 2013;14(1). doi:10.1186/1471-2164-14-666
29. Glastad KM, Hunt BG, Yi S V., Goodisman MAD. DNA methylation in insects: On the brink of the epigenomic era. *Insect Mol Biol*. 2011;20(5):553-565. doi:10.1111/j.1365-2583.2011.01092.x
30. Glastad KM, Hunt BG, Goodisman MA. Evolutionary insights into DNA methylation in insects. *Curr Opin Insect Sci*. 2014;1:25-30. doi:10.1016/j.cois.2014.04.001
31. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biol*. 2010;8(11). doi:10.1371/journal.pbio.1000506
32. Bonasio R, Li Q, Lian J, et al. Genome-wide and caste-specific DNA methylomes of the ants camponotus floridanus and harpegnathos saltator. *Curr Biol*. 2012;22(19):1755-1764. doi:10.1016/j.cub.2012.07.042
33. Terrapon N, Li C, Robertson HM, et al. Molecular traces of alternative social organization in a termite genome. *Nat Commun*. 2014;5(May). doi:10.1038/ncomms4636
34. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science (80-)*. 2010;328(5980):916-919. doi:10.1126/science.1186366
35. Werren JH, Baldo L, Clark ME. Wolbachia: Master manipulators of invertebrate biology. *Nat Rev Microbiol*. 2008;6(10):741-751. doi:10.1038/nrmicro1969

36. Zhang G, Hussain M, O'Neill SL, Asgari S. Wolbachia uses a host microRNA to regulate transcripts of a methyltransferase, contributing to dengue virus inhibition in Aedes aegypti. *Proc Natl Acad Sci*. 2013;110(25):10276-10281. doi:10.1073/pnas.1303603110
37. Vannier N, Mony C, Bittebière A-K, Vandenkoornhuyse P. Epigenetic Mechanisms and Microbiota as a Toolbox for Plant Phenotypic Adjustment to Environment. *Front Plant Sci*. 2015;6(December):1-8. doi:10.3389/fpls.2015.01159
38. Armaleo D, Miao V. Symbiosis and DNA methylation in the Cladonia lichen fungus. *Symbiosis*. 1999;26(2):143-163.
39. Li Y, Liew YJ, Cui G, et al. DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model Aiptasia. *DoiOrg*. Published online 2017:213066. doi:10.1101/213066
40. Takahashi K, Sugi Y, Nakano K, et al. Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells. *J Biol Chem*. 2011;286(41):35755-35762. doi:10.1074/jbc.M111.271007
41. Yu D-H, Gadkari M, Zhou Q, et al. Postnatal epigenetic regulation of intestinal stem cells requires DNA methylation and is guided by the microbiome. *Genome Biol*. 2015;16(1):211. doi:10.1186/s13059-015-0763-5
42. Negri I, Franchini A, Gonella E, et al. Unravelling the Wolbachia evolutionary role: the reprogramming of the host genomic imprinting. *Proc R Soc B Biol Sci*. 2009;276(1666):2485-2491. doi:10.1098/rspb.2009.0324
43. Douglas AE. The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. *Curr Opin Insect Sci*. 2017;23:65-69. doi:10.1016/j.cois.2017.07.012
44. Serbus LR, Rodriguez Garcia B, Sharmin Z, Momtaz AJMZ, Christensen S. Predictive Genomic Analyses Inform the Basis for Vitamin Metabolism and Provisioning in Bacteria-Arthropod Endosymbioses. *Genes|Genomes|Genetics*. 2017;7(6):1887-1898. doi:10.1534/g3.117.042184
45. Marhold J, Rothe N, Pauli A, et al. Conservation of DNA methylation in dipteran insects. *Insect Mol Biol*. 2004;13(2):117-123. doi:10.1111/j.0962-1075.2004.00466.x
46. Boffelli D, Takayama S, Martin DIK. Now you see it: Genome methylation makes a comeback in drosophila. *BioEssays*. 2014;36(12):1138-1144. doi:10.1002/bies.201400097
47. Gowher H, Leismann O, Jeltsch a. DNA of Drosophila melanogaster con- tains 5-methylcytosine. *Embo*. 2000;19(24):6918–6923.
48. Lyko F, Ramsahoye BH, Jaenisch R. DNA methylation in Drosophila melanogaster. *Nature*. 2000;408:538–540. doi:doi:10.1038/35046205
49. Chen T, Li E. Structure and Function of Eukaryotic DNA Methyltransferases. *Curr Top Dev Biol*. 2004;60:55-89. doi:10.1016/S0070-2153(04)60003-2
50. Kunert N, Marhold J, Stanke J, Stach D, Lyko F. A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development*. 2003;130(21):5083-5090. doi:10.1242/dev.00716
51. Goll MG, Kirpekar F, Maggert KA, et al. Methylation of tRNA Asp by the DNA Methyltransferase Homolog Dnmt2. *Science*. 2006;311(January):395-398. doi:10.1126/science.1120976
52. Waalwijk C, Flavell RA. MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaii sites. *Nucleic Acids Res*. 1978;5(9):3231-3236. doi:10.1093/nar/5.9.3231

53. Castellani A. On the discovery of a species of trypanosoma in the cerebrospinal fluid of cases of sleeping sickness. *Lancet*. 1903;161(4164):1735-1736. doi:https://doi.org/10.1016/S0140-6736(01)70338-8
54. Lindenbaum J, Klipstein FA. Folic acid clearances and basal serum folate levels in patients with thyroid disease. *J Clin Pathol*. 1964;17(6):666-670.
55. Greiner PO, Zittoun JM, Cheron JM. Pharmacokinetics of Valproic Acid After Oral and Intravenous Administration. *Br J Clin Pharmacol*. 1978;5(4):313-318. doi:10.1111/j.1365-2125.1978.tb01713.x
56. Pietrzik K, Bailey L, Shane B. Folie Acid and L-5-Methyltetrahydrofolate Comparison of Clinical Pharmacokinetics and Pharmacodynamics. 2010;49(8):535-548.
57. Aksoy S. Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *Int J Syst Evol Microbiol*. 1995;45:848-851. doi:10.1099/00207713-45-4-848
58. Chen X, Li S, Aksoy S. Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *J Mol Evol*. 1999;48(1):49-58. doi:10.1007/PL00006444
59. Balmand S, Lohs C, Aksoy S, Heddi A. Tissue distribution and transmission routes for the tsetse fly endosymbionts. *J Invertebr Pathol*. 2013;112(SUPPL.1):116-122. doi:10.1016/j.jip.2012.04.002
60. 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 Evol Microbiol*. 1999;49:267-275. doi:10.1099/00207713-49-1-267
61. Alam U, Medlock J, Brelsfoard C, et al. Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the Tsetse fly glossina morsitans. *PLoS Pathog*. 2011;7(12). doi:10.1371/journal.ppat.1002415
62. International Glossina Genome Initiative. Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis. *Science (80-)*. 2014;344(6182):380-386. doi:10.1126/science.1249656
63. Akman L, Yamashita A, Watanabe H, et al. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nat Genet*. 2002;32(3):402-407. doi:10.1038/ng986
64. Rio RVM, Symula RE, Wang J. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont Wigglesworthia. *MBio*. 2012;3(1):e00240-11. doi:10.1128/mBio.00240-11
65. Rio RVM, Attardo GM, Weiss BL. Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy. *Trends Parasitol*. 2016;32(9):739-749. doi:10.1016/j.pt.2016.05.002
66. Snyder AK, Rio RVM. “Wigglesworthia morsitans” folate (vitamin B<inf>9</inf>) biosynthesis contributes to tsetse host fitness. *Appl Environ Microbiol*. 2015;81(16):5375-5386. doi:10.1128/AEM.00553-15
67. Weiss BL, Wang J, Aksoy S. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol*. 2011;9(5):e1000619. doi:10.1371/journal.pbio.1000619
68. Bing X, Attardo GM, Vigneron A, et al. Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia : transcriptomic and metabolomic landscapes

reveal highly integrated physiological networks. *Proc R Soc B Biol Sci*. 2017;284(1857):20170360. doi:10.1098/rspb.2017.0360
69. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA methylation across insects. *Mol Biol Evol*. 2017;34(3):654-665. doi:10.1093/molbev/msw264
70. Nunes-da-Fonseca R, Berni M, Tobias-Santos V, Pane A, Araujo HM. Rhodnius prolixus: From classical physiology to modern developmental biology. *Toxicology*. 2009;23(4):1-14. doi:10.1177/0192513X12437708
71. Harington JS. Synthesis of thiamine and folic acid by Nocardia rhodnii, the Micro-symbiont of rhodnius prolixus. *Nature*. 1960;188(4755):1027-1028. doi:10.1038/1881027a0
72. Díaz S, Villavicencio B, Correia N, Costa J, Haag KL. Triatomine bugs, their microbiota and Trypanosoma cruzi: Asymmetric responses of bacteria to an infected blood meal. *Parasites and Vectors*. 2016;9(1):1-11. doi:10.1186/s13071-016-1926-2
73. Beard C Ben, Cordon-rosales C, Durvasula R V. Bacterial Symbionts of the T Riatominae and Their Potential Use in C Ontrol of Chagas Disease Transmission. *Annu Rev Entomol*. 2002;47:123-141.
74. Mesquita RD, Vionette- RJ, Lowenberger C, et al. Correction for Mesquita et al., Genome of *Rhodnius prolixus* , an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci*. 2016;113(10):E1415-E1416. doi:10.1073/pnas.1600205113
75. Walsh TK, Brisson JA, Robertson HM, et al. A functional DNA methylation system in the pea aphid, Acyrthosiphon pisum. *Insect Mol Biol*. 2010;19(SUPPL. 2):215-228. doi:10.1111/j.1365-2583.2009.00974.x
76. Vidal NM, Grazziotin AL, Iyer LM, Aravind L, Venancio TM. Transcription factors, chromatin proteins and the diversification of Hemiptera. *Insect Biochem Mol Biol*. 2016;69:1-13. doi:10.1016/j.ibmb.2015.07.001
77. Alvarenga EM, Mondin M, Martins JA, et al. Spatial distribution of AT- and GC-rich DNA within interphase cell nuclei of Triatoma infestans Klug. *Micron*. 2011;42(6):568-578. doi:10.1016/j.micron.2011.02.002
78. Reinhardt K, Siva-Jothy MT. Biology of the Bed Bugs (Cimicidae). *Annu Rev Entomol*. 2007;52(1):351-374. doi:10.1146/annurev.ento.52.040306.133913
79. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. Wolbachia as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci*. 2010;107(2):769-774. doi:10.1073/pnas.0911476107
80. Moriyama M, Nikohb N, Hosokawa T, Fukatsua T, McFall-Ngai MJ. Riboflavin Provisioning Underlies Wolbachia ' s Fitness Contribution to. *MBio*. 2015;6(6):e01732-15. doi:10.1128/mBio.01732-15.Editor
81. Medvedev SG. Classification of the order of fleas (Siphonaptera) and its theoretical foundations. *Entomol Obozr*. 1998;78(9):1080-1093. doi:10.1177/0192513X12437708
82. Iannino F, Sulli N, Maitino A, Pascucci I, Pampiglione G, Salucci S. Fleas of dog and cat: species, biology and flea-borne diseases. *Vet Ital*. 2017;53((4)):277-288. doi:doi: 10.12834/VetIt.109.303.
83. Erickson DL, Anderson NE, Cromar LM, Jolley A. Bacterial communities associated with flea vectors of plague. *J Med Entomol*. 2009;46(6):1532-1536. doi:10.1603/033.046.0642

84. Jones RT, Borchert J, Eisen R, MacMillan K, Boegler K, Gage KL. Flea-associated bacterial communities across an environmental transect in a plague-endemic region of Uganda. *PLoS One*. 2015;10(10):1-14. doi:10.1371/journal.pone.0141057
85. Lawrence AL, Hii SF, Chong R, et al. Evaluation of the bacterial microbiome of two flea species using different DNA-isolation techniques provides insights into flea host ecology. *FEMS Microbiol Ecol*. 2015;91(12):1-11. doi:10.1093/femsec/fiv134
86. Driscoll TP, Verhoeve VI, Gillespie JJ, et al. A chromosome-level assembly of the cat flea genome uncovers rampant gene duplication and genome size plasticity. *BMC Biol*. 2020;18(1):1-19. doi:10.1186/s12915-020-00802-7
87. Veracx A, Raoult D. Biology and genetics of human head and body lice. *Trends Parasitol*. 2012;28(12):563-571. doi:10.1016/j.pt.2012.09.003
88. Bonilla DL, Durden LA, Eremeeva ME, Dasch GA. The Biology and Taxonomy of Head and Body Lice-Implications for Louse-Borne Disease Prevention. *PLoS Pathog*. 2013;9(11). doi:10.1371/journal.ppat.1003724
89. Sasaki-Fukatsu K, Koga R, Nikoh N, et al. Symbiotic bacteria associated with stomach discs of human lice. *Appl Environ Microbiol*. 2006;72(11):7349-7352. doi:10.1128/AEM.01429-06
90. Braig HR, Perotti MA, Clark JM, et al. Correction for Kirkness et al., Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci*. 2011;108(15):6335-6335. doi:10.1073/pnas.1103909108
91. Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. Solving a bloody mess: B-vitamin independentmetabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the Leech haementeria officinalis. *Genome Biol Evol*. 2015;7(10):2871-2884. doi:10.1093/gbe/evv188
92. Clemons A, Haugen M, Flannery E, et al. Aedes aegypti: An emerging model for vector mosquito development. *Cold Spring Harb Protoc*. 2010;5(10). doi:10.1101/pdb.emo141
93. Ye YH, Woolfit M, Huttley GA, et al. Infection with a Virulent Strain of Wolbachia Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito Aedes aegypti. *PLoS One*. 2013;8(6). doi:10.1371/journal.pone.0066482
94. Nene V, Wortman JR, Lawson D, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*. 2007;316(June):1718-1723. doi:10.1126/science.1138878
95. David MR, Santos LMB Dos, Vicente ACP, Maciel-de-Freitas R. Effects of environment, dietary regime and ageing on the dengue vector microbiota: Evidence of a core microbiota throughout Aedes aegypti lifespan. *Mem Inst Oswaldo Cruz*. 2016;111(9):577-587. doi:10.1590/0074-02760160238
96. Holt RA, Subramanian GM, Halpern A, et al. The Genome Sequence of the Malaria Mosquito Anopheles gambiae. *October*. 2002;298(October).
97. Bianchi NO, Vidal-Rioja L, Cleaver JE. Direct visualization of the sites of DNA methylation in human, and mosquito chromosomes. *Chromosoma*. 1986;94(5):362-366. http://www.ncbi.nlm.nih.gov/pubmed/2881739
98. Oppold A, Kreß A, Vanden Bussche J, et al. Epigenetic alterations and decreasing insecticide sensitivity of the Asian tiger mosquito Aedes albopictus. *Ecotoxicol Environ Saf*. 2015;122:45-53. doi:10.1016/j.ecoenv.2015.06.036

99. Anderson JF, Magnarelli LA. Biology of Ticks. *Infect Dis Clin North Am*. 2008;22(2):195-215. doi:10.1016/j.idc.2007.12.006
100. Estrada-Peña A. Ticks as vectors: taxonomy, biology and ecology. *Rev Sci Tech l'OIE*. 2015;34(1):53-65. doi:10.20506/rst.34.1.2345
101. Bonnet SI, Binetruy F, Hernández-Jarguín AM, Duron O. The Tick Microbiome: Why Non-pathogenic Microorganisms Matter in Tick Biology and Pathogen Transmission. *Front Cell Infect Microbiol*. 2017;7(June):1-14. doi:10.3389/fcimb.2017.00236
102. Varela-Stokes AS, Park SH, Kim SA, Ricke SC. Microbial Communities in North American Ixodid Ticks of Veterinary and Medical Importance. *Front Vet Sci*. 2017;4(October):1-7. doi:10.3389/fvets.2017.00179
103. Clayton AL, Oakeson KF, Gutin M, et al. A Novel Human-Infection-Derived Bacterium Provides Insights into the Evolutionary Origins of Mutualistic Insect-Bacterial Symbioses. *PLoS Genet*. 2012;8(11). doi:10.1371/journal.pgen.1002990
104. Zhong J, Jasinskas A, Barbour AG. Antibiotic treatment of the tick vector Amblyomma americanum reduced reproductive fitness. *PLoS One*. 2007;2(5):1-7. doi:10.1371/journal.pone.0000405
105. Smith TA, Driscoll T, Gillespie JJ, Raghavan R. A Coxiella-like endosymbiontis a potential vitamin source for the lone star tick. *Genome Biol Evol*. 2015;7(3):831-838. doi:10.1093/gbe/evv016
106. Hunter DJ, Torkelson JL, Bodnar J, et al. The rickettsia endosymbiont of ixodes pacificus contains all the genes of de novo folate biosynthesis. *PLoS One*. 2015;10(12):1-15. doi:10.1371/journal.pone.0144552
107. Driscoll TP, Verhoeve VI, Guillotte ML, et al. Wholly rickettsia! reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells. *MBio*. 2017;8(5). doi:10.1128/mBio.00859-17
108. Gulia-Nuss M, Nuss AB, Meyer JM, et al. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nat Commun*. 2016;7(May 2015). doi:10.1038/ncomms10507
109. Meyer JM, Kurtti TJ, Van Zee JP, Hill CA. Genome organization of major tandem repeats in the hard tick, Ixodes scapularis. *Chromosom Res*. 2010;18(3):357-370. doi:10.1007/s10577-010-9120-4
110. Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. *Genome Biol Evol*. 2018;(March). doi:10.1093/gbe/evy066/4943971
111. Dadd RH. Insect Nutrition: Current Developments and Metabolic Implications. *Annu Rev Entomol*. 1973;18(1):381-420. doi:10.1146/annurev.en.18.010173.002121

CHAPTER 5: General Conclusions

The field of insect symbiosis is rapidly evolving. Understanding of the microbe-host interplay at the different levels, such as biochemical, genetic, and physiological, gives unique insights into the evolution of associations between such disparate organism like insects and bacteria. Similarly, the mechanisms operating at organismal scales facilitate the explanation of host adaptations that would be otherwise unlikely to occur, for example, the feeding habits of insects with restricted diets such as plant sap or vertebrate blood.

A combination of distinct approaches proves the best for the dissection of the mechanism regulating the evolution and interaction of symbiotic partners. A targeted approach with the tools of microbiology and entomology where mutant bacterial strains are inoculated into insect systems allows gathering evidence supporting the involvement of specific pathways towards the establishment and persistence of a bacterial-insect symbiosis, such as the case of the tsetse fly species and their bacterial partner *Sodalis*. Furthermore, recently developed technology for genomic studies, such as next-generation sequencing, allows an unbiased exploration of transcriptomic landscapes aiming at the discovery of profiles that may help to explain distinctive host traits, such as how different tissues have different permissiveness towards bacterial establishment or parasite invasion. Such high-throughput approaches are particularly advantageous to the study of symbioses, as they may be directed to the simultaneous analysis of both host and microbial profiles. For example, they may shine light on how the combination of tsetse immunity genes expression and *Sodalis* metabolism likely combine to explain differential susceptibility toward trypanosome infection in two tsetse species of different vector competence. Lastly, there is already an enormous amount of knowledge on insect physiology, vector biology and ecology, and bacterial metabolism, whose research may be exceedingly fruitful when attempting to explain mechanisms mediating adaptations to symbiotic relationships. For example, how vitamin provisioning has the potential to be harnessed by microbes to influence insect host gene transcription via epigenetic mechanisms. Put together, these different approaches will provide a necessarily nuanced framework for the understanding of insect symbioses in general and the tsetse holobiont in particular. A summary of the different interactions occurring between symbionts within the tsetse model is presented in Fig. 26.

Besides fulfilling the need for basic biological knowledge on insect symbioses, this research has the potential to deliver practical tools for epidemiology. The approaches outlined above allow the identification of targets and tools for vector-control strategies. For example, the identification of an ideal bacterial candidate for paratransgenesis, for which *S. praecaptivus* shows great potential. Equally helpful is the identification of genes or pathways that can be altered in order to disrupt microbial interactions, such as quorum sensing in bacteria, or immunity in insects, aiming at the modification of a host trait. For example, decreasing vector population, modifying the vector refractoriness towards a pathogen, or diminishing vector ability to transmit the pathogen.

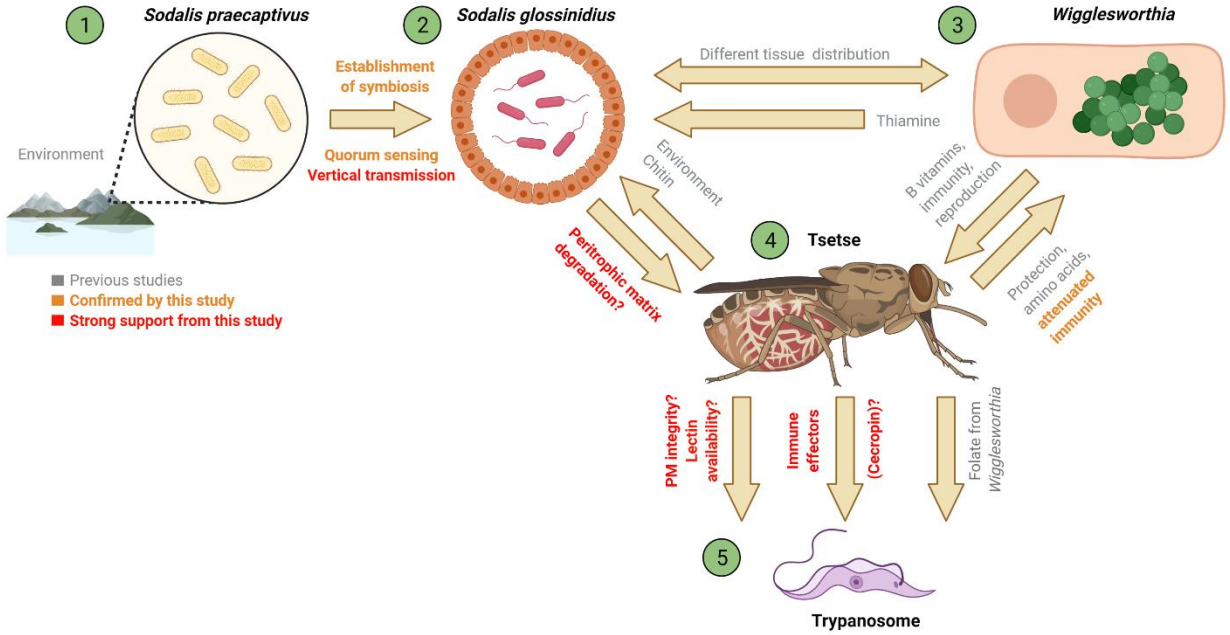

Figure 26. Tsetse symbiosis and vector competence.

① There is an environmental source for the bacteria that will take the evolutionary leap of becoming obligate symbionts of blood feeding insects, highlighted here are the *Sodalis*-like progenitors represented by *S. praecaptivus*. Physiological mechanisms such as quorum sensing and adaptation to vertical transmission facilitate this transition. ② Within tsetse, *S. glossinidius* benefits from a stable environment, low competition, and a rich nutrient source. In turn, the bacterium is hypothesized to have a detrimental impact on the structural integrity of the peritrophic matrix and trypanosome-fighting abilities of the fly. *S. glossinidius* is excluded from the bacteriome, but still capitalizes on resources produced by *Wigglesworthia*, such as thiamine. ③ Harboring a mutualistic symbiont in a dedicated and permissive organ facilitates the metabolic exchange with the host, which involves the provisioning of B vitamins form *Wigglesworthia* and amino acids from the fly. ④ The fly is serving as the microenvironment for interactions between symbionts as well as having an active role in which the fly influences the symbiont composition, abundance and their segregation to discrete tissues via processes such as vertical transmission and immunity. ⑤ Besides the tsetse biology, beneficial and detrimental interactions within the tsetse holobiont have an aggregate effect on the susceptibility of the fly toward trypanosome acquisition, development, and transmission, which translates on an impact on vector competence (Created with BioRender.com).

END