WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2013

Microbial Genome Evolution Due to Multifaceted Symbiosis within the Tsetse Fly (Diptera: Glossinidae)

Anna Kathleen Snyder
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation

Snyder, Anna Kathleen, "Microbial Genome Evolution Due to Multifaceted Symbiosis within the Tsetse Fly (Diptera: Glossinidae)" (2013). *Graduate Theses, Dissertations, and Problem Reports*. 544.
https://researchrepository.wvu.edu/etd/544

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Microbial Genome Evolution Due to Multifaceted Symbiosis within the Tsetse Fly (Diptera: Glossinidae)

Anna Kathleen Snyder

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

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Biology

Rita V. M. Rio, Ph.D., Chair
Serap Aksoy, Ph.D.
Nyles Charon, Ph.D.
Letha Sooter, Ph.D.
Jeffrey D. Wells, Ph.D.

Department of Biology

Morgantown, West Virginia
2013

Keywords: symbiosis, evolution, holobiont, tsetse, *Sodalis*, *Wigglesworthia*
Copyright 2013 Anna K. Snyder

ABSTRACT

Microbial Genome Evolution Due to Multifaceted Symbiosis within the Tsetse Fly (Diptera: Glossinidae)
Anna K. Snyder

Microbes are capable of rapid genetic modification, enabling the habitation of a wide field of niches, including forming interdependent associations with macroscopic hosts. While ancient multipartite mutualisms have been shown to involve metabolic complementation, little is known concerning the early genomic adaptations leading towards co-residence within a novel host. The overall objective of this research is to gain insight on genome evolution resulting from symbiosis, particularly by examining bacteria with varying levels of host dependency and times of establishment. The tsetse fly (Diptera: Glossinidae) serves as a relatively simple model system to investigate evolutionary aspects of symbiosis, while also maintaining medical and agricultural significance as vectors of African trypanosomes. In addition to potentially harboring trypanosomes, the tsetse enteric microbiota consists of two γ-Proteobacteria: the anciently associated obligate mutualist *Wigglesworthia* spp. and the recently established commensal *Sodalis glossinidius*. The genomes of *Wigglesworthia* spp. (isolated from *Glossina morsitans* (Wgm) and *G. brevipalpis* (Wgb)), *Sodalis* and *Trypanosoma brucei* subspp. have been sequenced and annotated, facilitating empirical studies exploring potential partner interactions and adaptations. My work first examines the importance of nutrient provisioning, specifically thiamine (Vitamin B1), for the maintenance of a stable symbiotic environment within the tsetse host. These studies demonstrated that *Sodalis* required exogenous thiamine for proliferation due to the erosion of biosynthetic capabilities, while *Wigglesworthia* thiamine biosynthetic loci expression was influenced by the functional demand for this nutrient. My research also explored how distinct symbiont metabolic capabilities, retained by Wgm, but lacking in the Wgb genome, contribute to host biology and phenotypic variation. Wgm chorismate and folate (Vitamin B9) biosynthesis increased during times of nutrient stress, such as pregnancy and trypanosome infection, and was found to be critical for host biology. Lastly, genetic adaptations leading towards symbiont diversification and establishment in novel hosts were investigated. To accomplish this, molecular phylogenetic analyses were performed on *Sodalis* and closely related bacteria using genome regions traditionally associated with accelerated evolution, such as surface encoding loci and internal transcribed spacer regions, further increasing the resolution of this clade. This enhanced knowledge of tsetse symbionts increases our understanding of tsetse biology, potentially contributing to disease control strategies, and offers additional insights regarding fundamental evolutionary aspects involved in microbial symbiosis.

Acknowledgements

I would first like to thank my doctoral advisor, Rita Rio, Ph.D., who pushed me beyond comfortable limits so that I could accomplish more than I ever imagined. The countless hours we spent together in the lab and in your office (reviewing, reading and rereading manuscripts) shaped me into a better scientist, taught me to persevere through obstacles and be my hardest critic. You also showed me throughout my graduate career to sometimes leave the lab and go to scientific meetings, be active in social groups, share my work in public settings and cherish hobbies. It is during these times when out-of-the-box ideas form, collaborations are made, and you are reminded of the value of your work. From the bottom of my heart, thank you.

I also want to thank my committee members, Letha, Nyles, Serap and Jeff. Your comments, suggestions and feedback were critical to making me the scientist I am today and elevated my research to the next level.

I next need to graciously acknowledge all of the individuals that contributed to my dissertation research. First, thank you to my collaborators in the Aksoy lab at Yale who made the trypanosome challenge studies possible. These experiments opened an avenue of my research that made me think beyond individual species interactions and better understand the interworking of this vector. Additionally, this dissertation would not have been possible without the many hours of lab work and discussions with past and present lab members; so thank you Brittany, Bidur, Vivian and the many undergraduates; specifically Ben, Colin, Cory, Derek, Britton, Lindsey, Cindy, Autumn, Zach, Pete and Brett. The time that I spent with you in lab doing anything and everything (from feeding tsetse and dissecting them, performing 96 DNA isolations in one afternoon, long days working with confocal imaging and analysis, and cell culture, to even watching hilarious parodies on YouTube) strengthened me as a researcher, mentor, teacher, problem solver and made me an overall better person.

I have found that graduate school is just as emotionally as it is mentally taxing. I'm very grateful for the friendships that I've made within the department and university. The unwavering support from fellow graduate students in the department, Shalaka, Mohna, Amy, Jessi Brie, Jen, Alyssa, and Stephanie, helped me through the high and lows. I also couldn't have done this without the Biology Department staff; Pat, Wendy, Kristi, Mickey, Diana and Judy. I am also sincerely thankful for the WVU Equestrian Team, who took me in as their own and gave me some work-life balance during the first years here.

Lastly, I need to thank my family and loved ones for their support throughout this journey. While I was MIA, spending long hours in the lab and office, your continued encouragement kept me going. Shelby, without you and Bella, I would not have eaten or smiled during my prelims. Last, but definitely not least, Luke, my love, you are the rock that keeps me grounded, the one who laughs when I’m unjustifiably upset, and are always telling me to believe in and stick up for myself. I don't know where I would be without you.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction*

Symbiosis: the importance of microbial associations for all life

Species interactions, across and within the domains of life, are ubiquitous in nature, fundamental in ecology and pivotal towards evolutionary diversification. Throughout history, metazoans have formed intimate partnerships with microorganisms, resulting in integral roles within host biology. These long-term, physical associations, termed symbioses, were first described in the 1870s (1, 2). Symbioses can lie within a gradient from mutualistic, where both partners benefit from the relationship, to parasitic, when one partner benefits at the cost of the other. These interactions can also range in their level of intimacy from obligate, essential for survival by one or both partners, to facultative, occurring if the opportunity presents itself. Since the initial descriptions of symbioses, much work in this field has greatly expanded our knowledge of the roles and importance microbes play in the biology of all living forms. For example, some of the many described functions of microbes include the fixation of nitrogen (3), providing nutrients lacking in the host diet (4), facilitating digestion (5, 6), contributing towards immune development and stimulation (7, 8), and aiding in the defense against pathogens (9, 10).

Symbiotic relationships are dynamic, thereby constantly evolving, adapting to changes in environmental and ecological conditions. Microorganisms or microbial communities, which are highly diverse and inhabit every facet of nature, play vital roles in the evolutionary success of a symbiosis, by more rapidly adapting to environmental changes through acquisition or modification of capabilities. For example, when stressed with drought conditions, plant fitness

* Adapted from: Snyder AK and Rio RVM. 2013. Interwoven biology of the tsetse holobiont. *J. Bacteriol*. 195: 4322-4220.

has been strongly linked to the rapid adaptations by the soil microbial community to the new environmental conditions (11). These microbial symbionts may also provide a source of evolutionary innovation, such as through the contribution of novel biochemical capabilities, potentially enabling their host to expand or change their ecological niche. Many associations are formed on the basis of nutrient provisioning, permitting the host to persist on a limited diet such as wood, plant phloem, or blood (4, 12). Additionally, during times of stress, such as increasing environmental temperatures (13-18), heightened predation (19), or limiting nutrient resources (5, 20), the formation of partnerships among species, particularly microbial, have resulted in enhanced fitness and ecological success. In return, the host may provide a protected niche for the microbes to colonize (such as bacteriomes or root nodules), often reducing inter- and intra-species competition, with the reciprocity of benefits contributing towards the maintenance of the relationship through successive generations. This theory is known as partner fidelity feedback, which describes how the evolution of symbioses through positive feedback, leading towards the enhanced fitness of partners, rather than punishing cheating behavior may be sufficient to maintain the mutualism through time (21). An alternative theory, known as host sanctions, entails the host punishing cheating behavior by symbionts that would ultimately harm the mutualism, thereby maintaining only those partners contributing to the association (21).

The holobiont concept

Due to the important role of symbionts towards host fitness, research has increasingly focused on a more holistic examination of the biological system, encompassing the host (animals and plants) and associated microbes, termed the holobiont (22) or the metaorganism (23), with the cumulative genetic material known as the hologenome (24). The hologenome theory of evolution has been used to examine the holobiont as a single unit undergoing evolution, adapting

to persist in or expand its niche (24, 25). This theory is based on four assumptions, the first being that all metazoans are associated with microbial symbionts (25, 26). Second, the fitness of the holobiont requires cooperation among the partners, as conflict may prove detrimental to overall health. In support, host immune systems have been shown to not only defend against pathogens, but also tolerate and possibly regulate symbiont populations, maintaining homeostasis (27-30). Third, the hologenome can change through alterations of the genetic material of any partner. Symbionts can allow the holobiont to adapt more quickly to ecological disturbances, through the acquisition of novel capabilities by horizontal gene transfer, or changes in population dynamics or community composition, thereby aiding in the persistence of the holobiont. For example, the bean bug, *Riptortus pedestris*, was recently shown to rapidly acquire insecticide resistance by the acquisition of strains of its symbiotic gut bacteria with enhanced insecticide-degrading capabilities from the soil (31). Lastly, the symbiotic associations must be passed on through generations of the host, maintaining the species composition of the holobiont. Many strategies have evolved to ensure the transmission of a host to its offspring, including vertical symbiont transmission in reproductive cells (32, 33), egg smearing (reviewed in (4)), and feeding of feces to offspring (for example baby koalas eating 'pap' (34)), or horizontal transmission, including symbiont acquisition through sea water (35, 36), or soil (37, 38). By applying the holobiont concept and hologenome theory of evolution to symbioses, research will elucidate novel mechanisms driving microbial species cooperation and adaptation, which enable successful co-occupancy within a specific niche.

Traditionally used in aquatic biology, applications of the holobiont concept have provided insights into marine microbiology such as coral health and the functioning of deep-sea hydrothermal vents (22, 24, 39). Due to the pivotal role of symbionts towards host fitness, the

concept of holobionts has recently extended into other facets of biology (23, 40). Heightened recognition of microbial symbionts as major contributors to host health has spurred the Human Microbiome Project, aimed at characterizing the microbial communities of several distinct spatial sites on the human body to better our understanding of their role in health and disease (41). These communities are highly complex (42) and can be composed of hundreds of species-level phylotypes, determined by ≥ 97% 16S rRNA sequence identity (43, 44). To fully understand processes occurring within complex holobionts, examination of more simple systems may aid in dissecting intimate interplay among the partners.

Insects as symbiosis model systems

First studied by Paul Buchner (4), insects provide naturally simple symbiosis models, enabling a more complete examination of both host-microbe and microbe-microbe interactions. Unlike the complex microbiota of mammals, many insects harbor low complexity microbiomes, with greater diversity and variation in the types of interactions. For example, unlike the acidic environment of mammalian stomachs, insects harbor a range of gut environments, from basic lepidopteron caterpillars to very acidic higher Dipteran, thereby increasing the diversity of microbes with which associations are formed (reviewed in (45)). Through comparative studies of multiple insects symbioses, broad scale mechanisms influencing the evolution of complex mutualisms and promoting coexistence (reviewed in (45)), as well as genomic effects of ancient associations (reviewed in (46)) have been elucidated. For example, the host immune system is now recognized in maintaining a symbiosis through interactions between bacterial symbionts, such as the up-regulation of specific loci in the presence of symbionts to control population density (28, 47, 48). Additionally, some associations are maintained by measures taken to prevent immune activation, such as avoidance by the symbiont through intracellular localization

or loss of immunity related genes by the host (45). Moreover, biotechnological advances, such as genomics (49-51), transcriptomics (52-54), proteomic analysis utilizing variations of mass spectrometry (55, 56), and '*in situ*' and live cell imaging (47), have collectively enabled a more in depth understanding of genomic complementation and metabolic interplay among partners.

The association of microbes within a symbiosis context may also drive specialization. The basis of many insect-microbial symbioses lies in nutrient provisioning, often supplementing a limited host diet. Genome sequencing of primary symbionts from aphids (57-59), tsetse flies (51, 60), sharpshooters (61, 62), and cicadas (63) provide strong support for their putative roles by encoding biosynthetic pathways for vitamins or amino acids lacking in the blood or plant limited host diet. Additionally, these genomic studies have provided insight into the molecular and evolutionary consequences of insect-bacterial mutualisms, revealing characteristic genomic traits occurring in light of a symbiotic lifestyle. For example, anciently associated bacterial endosymbionts encode highly reduced, adenine-thymine (A-T) rich genomes, with a deletion of loci not required for the maintenance of the symbiosis, including metabolic pathways not essential in the constant intracellular host environment (reviewed in (46, 64)). These genome changes are brought about by the collective action of high genetic drift, resulting from extreme bottlenecks during vertical transmission and small effective population sizes within hosts (65), along with the relaxed selection of non-essential genes and redundant pathways. Additional evidence for genome reduction is that recently established bacteria, still undergoing the transition from free-living to symbiotic lifestyle, encode a high number of pseudogenes (66, 67) that will most likely be purged as the association ages. These genomic characteristics of bacteria within mutualisms reflect the functional integration and complementation with the biology of their specific insect host.

Much of the pioneering work enhancing our understanding of insect host-microbe interactions has focused on the aphid, which feeds exclusively on plant phloem; a diet lacking in essential amino acids. The aphid maintains an ancient obligate mutualism with the γ-Proteobacterium *Buchnera aphidicola*, whose genome encodes the essential amino acid biosynthetic pathways utilized to supplement the restricted host phloem diet (57-59). Synchronous partner biology reflects the importance of the nutrient provisioning role by the symbiont towards the success of the holobiont. Early transcriptome analysis of the aphid bacteriocytes, which are at the symbiotic interface, resulted in the identification of amino acid transporter ESTs within this organ (52). Interestingly, some of the most abundant transcripts found in the study were characterized as invertebrate-type lysozymes, believed to break down bacterial cell walls, suggesting either a role in warding off unwanted transient microbes or aiding in the host tolerance of *Buchnera* (52). As determined by metabolic modeling, the production of essential amino acids by *Buchnera* may be strongly influenced by the host through the supply of metabolic precursors to its symbiont (68). Recently, proteomic analysis further deepened our understanding of this particular host-microbe interaction. While there was no evidence for selective protein transfer among the aphid host and *Buchnera*, results supported previous hypotheses of bacteriocyte transporters being utilized for metabolic transfer and further evidence for the specific production of host metabolic precursors by their corresponding enzyme enrichment in the bacteriocyte proteome (55).

In addition to harboring *Buchnera*, aphids have also been found in association with multiple facultative symbionts, such as *Serratia symbiotica, Hamiltonella defensa* (itself harboring a toxin-encoding phage protecting against parasitoid attacks) and *Regiella insecticola,* which have been shown to contribute novel capabilities to the aphid holobiont, including defense

against heat stress and parasites and an expansion of host-plant range (reviewed in (69)). Additionally, in the cedar aphid, *Cinara cedri*, the loss of biosynthetic loci by *Buchnera* appears to be complemented by the facultative symbiont *Serratia symbiotica*, resulting in shared roles of tryptophan and riboflavin biosynthesis by the bacterial partners (70, 71).

Culture independent techniques have enabled examination of slightly more complex model systems, demonstrating ways in which microbial communities work in concert within a host, by identifying genomic contributions arising from each partner. Similar to the cedar aphid symbiosis, the presence of multiple ancient microbial symbionts has been shown to result in complementary genome reduction, retaining unique biosynthetic capabilities required for the symbiosis. For example, the glassy winged sharpshooter (GWSS), *Homalodisca coagulata*, whose diet consists of amino acid deficient plant xylem, harbors two microbial partners, the *Bacteroidetes* species *Sulcia muelleri* (62, 72) and the γ-Proteobacterium *Baumannia cicadellinicola* (61). Comparative genomics revealed that *Sulcia* provisions most of the essential amino acids, while *Baumannia* supplies the remaining amino acids and additional vitamins and cofactors (61, 62), demonstrating extensive genetic complementation between the symbiont species. The cicada, *Diceroprocta semicincta*, also houses *S. muelleri*, with very similar genetic composition (63) to that of the GWSS symbiont. Interestingly, its co-resident symbiont, the α-Proteobacterium *Hodgkinia cicadicola*, has undergone convergent evolution with *Baumannia*, encoding the amino acid biosynthetic capabilities lacking in *Sulcia* (63). A similar example has also been described in the spittlebug, *Clastoptera arizonana*, which also houses *S. muelleri*, in addition to the β -Proteobacterium *Candidatus* Zinderia insecticola (73). These symbionts have perfectly complementary amino acid biosynthesis capabilities within its spittlebug host (73). An even more intimate example is that of the mealybug, *Planococcus citri*, which harbors dual

ancient symbionts, the γ-Proteobacterium *Candidatus* Moranella endobia, which lives inside the β-Proteobacterium *Candidatus* Tremblaya princeps (74). The genomes of both of these symbionts were recently annotated revealing little overlap in content, yet neither bacterium encodes a complete amino acid biosynthetic pathway. Instead, the synthesis of essential amino acids was shown to necessitate a medley of gene products arising from both bacterial partners, and possibly the participation of the mealybug host for completion (49). The observed metabolic interdependency of these symbiotic systems, generated through complementary genome evolution, maintains the necessity of microbial partners, potentially preventing antagonism among them through the division of labor, along with increasing the efficiency of the holobiont.

The tsetse fly

Localized exclusively to Sub-Saharan Africa, there are approximately 31 species and subspecies of tsetse flies (Diptera: Glossinidae), which can be divided into 3 groups; *morsitans*, *palpalis*, and *fusca* (75). Tsetse species also differ in their range of habitats, from rain forests to woodlands and savannas, and have also been shown to have different blood meal host animal preferences (76-78). The unique biology of tsetse contributes to the maintenance of a simple larval microbial community, consisting of only 3 maternally transmitted bacterial symbionts (79), in comparison to other Diptera, such as mosquitoes and fruit flies, which harbor a greater complexity of bacterial taxa (80, 81). One distinct feature of tsetse biology is that both sexes maintain a strictly hematophagous lifestyle, persisting solely on vertebrate blood, which limits the introduction of additional microbes through an oral/digestive route. Due to their restricted diet, tsetse rely on microbial symbionts for provisioning essential metabolites lacking in blood (12). Another contributing factor to microbiome simplicity is the tsetse reproductive strategy, known as adenotrophic viviparity. This reproductive strategy involves high maternal investment

and a low reproductive output of only 6-8 offspring in their 3-4 month lifespan (75). Unlike many higher Dipteran, female tsetse have highly modified reproductive tracts (82), enabling the deposition of a single fertilized egg into a muscular uterus, which is connected to highly specialized accessory glands, referred to as milk glands. Milk secretions provide nourishment and a route through which microbial symbionts (83, 84) are transferred during intrauterine larval development. This form of reproduction transmits the microbiota with high fidelity, while preventing exposure to transient microbes during early tsetse development (79).

In addition to its use as a model system for understanding the evolution of a holobiont, tsetse maintains significance as the sole and obligate vector of the protozoan African trypanosomes (*Trypanosoma spp*.). These parasites (*T. brucei rhodesiense* and *T. b. gambiense*) are the causative agents of Human African Trypanosomiasis (HAT; commonly called 'sleeping sickness'), a disease affecting the central nervous system that is lethal if left untreated. HAT threatens millions of people in approximately 36 countries and has been classified by Doctors Without Borders as a neglected tropical disease, impacting some of the poorest rural areas in Africa (97). Another African trypanosome, *T. b. brucei*, causes Nagana, a similar wasting disease in domesticated animals, particularly cattle, further impeding the economic development of affected areas (98). Disease relief is relatively non-existent as there are no vaccines available to prevent African trypanosomiasis, manual trapping is often unreliable due to the social unrest in many affected areas, diagnostics are limited, and the small arsenal of drugs available for treatment are associated with significant toxic side effects. Therefore, vector control is an alternative intervention to break the disease cycle (99), as advocated with other systems (reviewed in (100)). For example, *Wolbachia* has been shown to inhibit the replication of multiple arboviruses and filarial nematodes within *Aedes* mosquitoes (101-104), as well as

shorten the host life span, not permitting cyclical pathogen development and transmission (105). Additionally, symbionts within the guts of mosquitos and triatome bugs are being genetically modified to produce anti-parasitic molecules, in efforts to block transmission of malaria (106) and Chagas disease (107), respectively. Knowledge on tsetse fly symbiosis not only stands to provide basic insight into how microbial partners adapt and respond to changes in ecological factors and parasite infections, but may also be of applied value to generate novel modes of pest biocontrol (100).

Tsetse microbial community

The tsetse microbiota consists primarily of three vertically transmitted bacterial species (Fig. 1). These microbes include two enteric γ-Proteobacteria, the obligate mutualist *Wigglesworthia spp.* (85), and the commensal, *Sodalis glossinidius* (86). Tsetse can also harbor the α-Proteobacteria *Wolbachia* (87), a facultative parasite infecting many different invertebrates (88, 89), which is typically restricted to the reproductive organs (90, 91). Field studies report a more complex diversity in adult flies (92-94), although these microbes are believed to be transient in nature. The tsetse holobiont provides opportunities to examine evolutionary aspects associated with adapting to microbial co-residence, as *Wigglesworthia* and *Sodalis* have drastically different times of establishment (95, 96). Moreover, interactions among microbes with varying levels of host dependency and symbiotic roles can be empirically investigated.

Wigglesworthia spp. maintain an obligate mutualism with tsetse and display significant concordant evolution with their specific host species, dating back 50-80 million years (95). In both sexes, this symbiont is localized intracellularly in specialized host cells (bacteriocytes) at the anterior midgut, collectively comprising an organ known as the bacteriome (Fig. 1). An additional extracellular population is found in the female milk glands, which are maternally

transmitted to offspring (83, 84). Described roles of this symbiont include both nutrient provisioning, where *Wigglesworthia* supplements B-vitamins lacking in the tsetse blood diet (12, 51, 60, 108), and contributions towards the maturation of host immunity (28, 79, 109). Insight into these roles has been found by examining tsetse biology upon removal of the symbiont. For example, the bacteriome population is vital for nutrient provisioning during host reproduction, as flies lacking the *Wigglesworthia* bacteriome populations are sterile (110-112), with fecundity partially restored by B-vitamin or yeast extract supplementation (12, 112). Absence of the milk gland symbiont population does not inhibit reproduction (111), as these are believed to be dedicated for vertical transmission and the persistence of the symbiosis through evolutionary time. In addition, the presence of *Wigglesworthia* during larval stages is essential for proper immune development, as larvae that lack this symbiont were significantly compromised in the induction of pathways associated with cellular immunity (79). The immuno-compromised phenotype of aposymbiotic larvae can be reversed by feeding their moms a diet supplemented with *Wigglesworthia* cell extracts (109). The presence of the tsetse's larval microbiota also contributes to the proper development of the adult peritrophic matrix, separating epithelial cells from the contents of the lumen, which regulates the timing of immune induction following parasite challenge (113). *Wigglesworthia* also impacts tsetse digestion, temperature sensitivity and susceptibility to infection with trypanosomes (28, 111, 113). It is important to mention that removing *Wigglesworthia* also causes indirect effects towards the host, not related to the loss of symbiont function. Some examples include additional nutritional deficiencies due to inhibited blood meal digestion (111), transient microbial colonization arising from an altered immune state (28, 79, 109), and perturbations to the remaining tsetse microbiota (114).

The annotation of *Wigglesworthia* genomes, isolated from *Glossina brevipalpis* (Wgb) (60) and *G. morsitans* (Wgm) (51), revealed characteristics similar to other ancient insect symbionts (46, 64, 115) including a reduced size (~0.7 Mb) with high adenine-thymine bias. Genome adaptations by *Wigglesworthia*, while in association with tsetse, are believed to have resulted in the loss of many capabilities required for a free-living lifestyle through reductive evolution (51, 60). Despite its small genome size, the majority of B-vitamin biosynthesis pathways remain intact, supporting the nutritional mutualism. Comparative analyses revealed that Wgm and Wgb maintain similar genomic repertoires with high synteny. Interestingly, some pockets of unique Wgm genes, potentially contributing to anabolic distinctions, were found (51). In Chapter 3, I examine the significance of the retention of the chorismate and downstream folate biosynthetic pathways by Wgm, lacking in Wgb, towards *G. morsitans* biology specifically towards life longevity, digestion and fecundity. As this symbiont has undergone deep co-diversification with the tsetse host (95), any unique capabilities by a *Wigglesworthia sp*. influencing physiological and phenotypic differences between host species remains largely unknown.

In contrast to the ancient *Wigglesworthia* association, tsetse's commensal partner, *Sodalis*, is believed to have established recently within the tsetse host from a previously free-living progenitor. Evidence lies in its ability to still be cultured (116) providing a tremendous benefit for empirical analyses, wide tsetse tissue tropism with both intra- and extra-cellular localization (84, 91), lack of co-evolution with host species (95), and stochastic presence in the field (92, 94, 117). Similar to *Wigglesworthia, Sodalis* is vertically transmitted through the maternal milk glands (83, 84). Despite a more recent association, *Sodalis* displays genomic signatures indicating that adaption to the symbiosis has commenced. While *Sodalis*' genome (4.2

Mb) is comparatively larger than *Wigglesworthia*'s, it appears to be undergoing massive reduction as it is composed of > 50% pseudogenes (66, 118), indicative of relaxed selection on non-essential genes. Notably, *Sodalis* has modified its outer membrane protein A (OmpA), which represents a molecular adaptation that contributes to host immune tolerance (119). The OmpA protein is also utilized in biofilm production within the tsetse gut, further protecting *Sodalis* from host immune responses (120).

Biotechnological advancements, notably culture-independent techniques, have accelerated the number of described host-associated bacteria, stimulating interest in examination of the Enterobacteriaceae *Sodalis*-allied clade, as members are present in a diverse array of insect and environmental samples (86, 121-132) (Table 1). While this bacterial group appears to have an enhanced ability to establish within a variety niches, relative to most other characterized symbionts, much of the initial molecular phylogenetic analyses within the *Sodalis*-allied clade has utilized the conserved 16S rRNA gene, resulting in low resolution (121, 126). A recently described member obtained from an environmental source, known as strain HS, has provided novel insights on the progenitor of this clade (132). Comparative genomic analyses of strain HS, with other members of the *Sodalis*-allied clade, specifically *Sodalis* and the *Sitophilus oryzae* primary symbiont (SOPE), revealed that both insect symbiont genomes were near-perfect subsets of the strain HS genome, yet each contained a unique set of pseudogenes (132). These results suggest that strain HS may be a representative environmental progenitor of the *Sodalis*–allied clade, which has independently formed symbioses with various insects. Continued examination of this clade will enhance knowledge of potential adaptations aiding in establishment within a broad range of niches, mechanisms facilitating host switching, and the impact of these host jumps towards symbiont genome evolution. Chapter 4 describes the efficacy of genome regions

traditionally associated with accelerated rates of evolution, to examine the divergence of the *Sodalis*-like symbiont clade and identify early genomic host-specific modifications, possibly aiding in adaptation and establishment within these insect hosts.

The third bacterial member of the tsetse holobiont is an intracellular pathogen within the genus *Wolbachia* (87). While there is a high prevalence of *Wolbachia* infections in lab colonies (133), field populations are more stochastic and infection is also not detected in all tsetse species (134). This symbiont is transovarially transmitted through successive host generations and has recently been shown to induce cytoplasmic incompatibility within the tsetse host (112), where developmental arrest of an embryo occurs when an infected male mates with an uninfected female. The association may also have a long co-evolutionary history with some tsetse species, as *Wolbachia* loci were found horizontally transferred into the host genome (134).

Impact of additional microbes on holobiont success

The introduction of additional microbes can influence the fitness of the holobiont. For example, salivary gland hypertrophy virus (SGHV) is a well-characterized parasitic infection in tsetse (135). SGHV, a nuclear rod-shaped, enveloped DNA virus (136), has low infection rates in the field and can be both vertically and horizontally transmitted. This infection can quickly spread in lab colonies, driven by horizontal transmission through artificial feeding systems, which can harbor concentrated viral numbers in the blood that would otherwise quickly disseminate within a vertebrate host (135). Viral infection is associated with testicular degeneration and ovarian abnormalities (137-139), which can lead to decreased tsetse fertility and longevity (140, 141). This is just one example of how parasitic associations may impact the tsetse host and influence evolutionary adaptations by the bacterial symbionts. The interactions of SGHV and the tsetse microbiota remain largely unknown. SGHV, as well as other potential

parasitic interactions, within tsetse populations should be considered when examining the success and evolution of the holobiont.

It should also be noted that recent field studies have found an unexpected diversity within the microbial community of tsetse, which is dependent on host species and geographic region (92, 94). Differences in abiotic conditions and food sources may influence the composition of these transient microbial communities (78, 117). Although additional microbes have been found in association with tsetse in the field, only *Wigglesworthia*, *Sodalis*, and *Wolbachia* are maternally transmitted, as 16S rRNA clone libraries of 3rd instar larvae only contain these 3 bacterial species (79).

Metabolic interactions among microbiota

While the genomic evolution and importance of individual symbiont species within tsetse has been examined, the community dynamics are only beginning to be explored. Although *Wigglesworthia* and *Sodalis* have different evolutionary histories with tsetse, they maintain parallel population dynamics through host development (142, 143), indicative of coordinated activities or a generalized level of host control. Unlike the extensive metabolic complementation observed within ancient co-resident symbionts (49, 61-63, 72, 73), comparative genomics reveals that *Sodalis* encodes a majority of *Wigglesworthia* genes (144). This genetic redundancy brings into question the factors contributing to the maintenance of both associations within the tsetse host and how cooperation, rather than competition, occurs among the symbionts. Co-localization of *Wigglesworthia* and *Sodalis* within the midgut may have led to the evolution of metabolic interactions. Within the tsetse holobiont, synergistic effects of co-residency were recently observed, as clearance of *Wigglesworthia* resulted in the loss of *Sodalis* over generations of the host (114), possibly due to metabolic dependencies, as previously suggested by similar trends in

population dynamics through tsetse development (142). Chapter 2 describes the metabolic interplay of thiamine monophosphate (a thiamine (B1) derivative) and its role in the maintenance of homeostasis within the tsetse holobiont, discussing how this nutrient is involved in symbiont population control and may play a role in preventing antagonism.

Understanding the tsetse holobiont for enhanced vector control

An additional major factor influencing the holobiont is trypanosome presence; as once a fly becomes infected with trypanosomes, they remain infected for the duration of their lifespan (Fig. 1). Tsetse flies play an obligate role in the successful development and transmission of *Trypanosoma spp.* (reviewed in (145)). A phenotypic difference between tsetse species (146-150) and sex (151-153) is their vector competency; i.e. their ability to support the development and transmission of trypanosomes to naïve hosts. Within the fly, *T. brucei subspp*. undergo many stages of developmental differentiation and multiple population bottlenecks. In vertebrate blood, *T. brucei subspp*. are found in two forms, a long slender bloodstream form, which will differentiate into a short, stumpy non-dividing forms when at high population densities, pre-adapted to survive in the tsetse midgut. The bloodstream form trypanosomes are taken up by tsetse in the blood meal and the parasites transform into a procyclics within the gut. After establishing within the midgut, *T. brucei subspp*. migrate into the ectoperitrophic space, possibly by crossing the peritrophic membrane (154, 155), enclosing the blood meal, and differentiate into mesocyclics. They then migrate anteriorly and invade the proventriculus, where they differentiate into long trypomastigotes, which then give rise to the long epimastigote form. The long epimastigote undergoes asymmetric division, generating long and short epimastigotes. The short epimastigotes migrate anteriorly and colonize the salivary glands, where they undergo genetic recombination (156) and a final differentiation, producing metacyclic forms, which can

then be transmitted to naïve hosts during tsetse feedings. While a tsetse fly may ingest an infected blood meal resulting in a midgut infection of trypanosomes, tsetse species differ in their ability to harbor mature infections of the salivary glands, which can be transmitted to naïve hosts (146-148, 153).

Complex mechanisms are involved in the susceptibility to trypanosome infection by tsetse. While in the fly, the parasites are also heavily bombarded by the tsetse immune system including the synthesis of antimicrobial peptides and the production of reactive oxygen species (157-161). There are also multiple population bottlenecks that *T. b. brucei subspp.* undergo before successfully establishing mature infections within the salivary glands (162). During the different stages of infection within tsetse, trypanosomes must persist through multiple rounds of attacks by the host immune system, including an initial attack of trypanocidal lectins, reactive oxygen intermediates and antimicrobial peptides, which impede successful establishment in the tsetse midgut (157), followed by additional immune responses from the proventriculus while migrating anteriorly through the foregut (158), and eventual colonizing of the salivary glands. Consequently, only a few trypanosomes serve as founders of the populations within each salivary gland (162). The presence of trypanosomes, as well as the related biological modifications within tsetse, such as heightened immune stimulation and increased competition for space and resources, may also impact their microbial symbionts.

Measures to prevent the spread of African trypanosomiasis have been historically targeted at controlling the tsetse fly population, such as using mass insecticide spraying during outbreaks and the release of sterile males in restricted areas to reduce field population sizes (145). Such measures have been successful within their targeted locales (163). Nevertheless, the threat of trypanosomiasis remains relevant as political instability (97) and decreased priority by

local authorities, due to the reduced numbers of reported cases (164), impedes the continuous efforts required to prevent the re-establishment and subsequent heightened disease incidence (165). By gaining a more holistic view of tsetse biology, insights into novel strategies for controlling the spread of the disease may be gained. In fact, a recent International Atomic Energy Agency (IAEA) coordinated research project (CRP) aims to unravel the interactions between the tsetse host, *Wigglesworthia*, *Sodalis*, *Wolbachia*, SGHV, and the development African trypanosomes to increase knowledge of ways to enhance refractoriness to trypanosome infection (166).

Past studies have demonstrated a positive correlation between *Sodalis* presence and trypanosome infections in field flies (117, 167). *Sodalis* is believed to contribute to the susceptibility of teneral flies (i.e. newly emerged unfed adults), through its endochitinase activity within the midgut which breaks down chitin (168), producing a by-product of N-acetyl-D-glucosamine which inhibits the action of trypanocidal lectins (168, 169).

The role of *Wigglesworthia* in the tsetse fly's susceptibility to trypanosome infection remains largely unknown. A link between *Wigglesworthia* and trypanosome infection was suggested as the removal of the symbiont resulted in higher susceptibility to midgut infection in older, non-teneral flies- a time point typically of low vector competency (111). Subsequently, the absence of *Wigglesworthia* was found to impair host immune system development (79). Thus, the higher susceptibility to trypanosome infection may be due to compromised immunity, although the effect of an altered nutritional state may also be a contributing factor. Genome comparisons between *Wigglesworthia spp.* revealed potential metabolome differences among the primary symbionts (51). One distinction lies in the complete retention of the chorismate (an intermediate in the production of aromatic compounds, including amino acids and vitamins) and

downstream folate (Vitamin B9) biosynthetic pathways by Wgm, but not Wgb. Interestingly, the parasitic lifestyle of *T. brucei subspp*. has resulted in a highly restricted genomic repertoire, compensating for the absence of biosynthetic pathways by encoding transporters to sequester metabolites (including folate) from the environment (170, 171). This enhanced biosynthetic capability may contribute to the higher reported vector competency of the *G. morsitans* host, relative to *G. brevipalpis* (146-148, 150), as trypanosomes necessitate exogenous folate for growth (172), and is further investigated and discussed in Chapter 3. Enhanced understanding of the unique capabilities of specific *Wigglesworthia spp.* towards holobiont functioning may contribute to a more holistic view of factors that result in different levels of refractoriness between tsetse species.

Tsetse immune tolerance of symbionts may also influence the fly's susceptibility to trypanosome infection. For example, the contribution of tsetse's immune system to the persistence of the *Wigglesworthia* symbiosis may also play a role in the fly's ability to transmit trypanosomes. The host pathogen recognition protein PGRP-LB, which scavenges peptidoglycan thus preventing immune deficiency (IMD) signaling pathway stimulation, is intimately associated with maintaining the *Wigglesworthia* symbiosis (28, 173). PGRP-LB is maternally transmitted via milk gland secretions to developing offspring and is only produced by adult flies after their first blood meal (173). This protein has also been shown to have anti-trypanosomal activity (173). Therefore, higher levels of PGRP-LB may aid in the refractory nature of older, non-teneral flies to trypanosome infection.

Field studies examining the rates of microbial co-infections within tsetse may also provide insight on symbiont interactions. One study examining the association of co-infections in *Glossina fuscipes fuscipes* in Uganda found a negative correlation between *Wolbachia* and

SGHV prevalence, while SGHV and trypanosome infection was positively correlated (174). This finding highlights the importance of examining the evolutionary and physiological effects of co-infection. One trypanosome control strategy that relies on symbiont interactions is known as paratransgenesis (for more detail see (99, 100, 175)). Paratransgenesis involves manipulating *Sodalis* to express anti-trypanosomal effector molecules, and utilizes the cytoplasmic incompatibility properties of *Wolbachia* (112, 176) to drive the genetically modified symbiont into natural populations.

Conclusions

The low complexity of the tsetse holobiont and the annotated genomes of its members enable investigations into the evolutionary aspects of co-residence and holobiont adaptations when challenged with both intrinsic and ecological disturbances. Comparisons of the tsetse holobiont, in which members are still transitioning into the symbiotic lifestyle, to other anciently co-evolved mutualisms will help describe mechanisms contributing to early establishment (Chapter 4), integration and cooperation within a microbial community, such as through metabolic interplay (Chapters 2 and 3). Moreover, a more holistic and comprehensive understanding of the tsetse holobiont may identify additional factors promoting or inhibiting vector competency that may ultimately aid in controlling the spread of African trypanosomiasis.

REFERENCES

1. **de Bary A.** 1879. Die Erscheinung der Symbiose. Verlag von Karl J. Trubner: Strassburg.
2. **Frank AB.** 1877. Über die biologischen Verhältnisse des Thallus einiger Krustflechten. Cohn Beitr Biol. Pflanz **2:**123–200.
3. **Virtanen AI.** 1945. Symbiotic nitrogen fixation. Nature **155:**747-748.
4. **Buchner** P. 1965. Endosymbiosis of animals with plant microorganisms. John Wiley & Sons, Inc. New York, NY.
5. **Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang XN, Hernandez M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450:**560-565.
6. **Mcallister TA, Bae HD, Jones GA, Cheng KJ.** 1994. Microbial attachment and feed digestion in the rumen. J. Anim. Sci. **72:**3004-3018.
7. **O'Hara AM, Shanahan F.** 2006. The gut flora as a forgotten organ. EMBO Rep. **7:**688-693.
8. **Ivanov II, Littman DR.** 2011. Modulation of immune homeostasis by commensal bacteria. Curr. Opin. Microbiol. **14:**106-114.

9. **Oliver KM, Degnan PH, Hunter MS, Moran NA.** 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science **325:**992-994.

10. **Ritchie KB.** 2006. Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. Mar. Ecol. Prog. Ser. **322:**1-14.

11. **Lau JA, Lennon JT.** 2012. Rapid responses of soil microorganisms improve plant fitness in novel environments. Proc. Natl. Acad. Sci. U.S.A. **109:**14058-14062.

12. **Nogge G.** 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82:**101-104.

13. **Montllor CB, Maxmen A, Purcell AH.** 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. Ecol. Entomol. **27:**189-195.

14. **Abrego D, Ulstrup KE, Willis BL, van Oppen MJ.** 2008. Species-specific interactions between algal endosymbionts and coral hosts define their bleaching response to heat and light stress. Proc. Biol. Sci. **275:**2273-2282.

15. **Burke G, Fiehn O, Moran N.** 2010. Effects of facultative symbionts and heat stress on the metabolome of pea aphids. ISME J. **4:**242-252.

16. **Littman RA, Bourne DG, Willis BL.** 2010. Responses of coral-associated bacterial communities to heat stress differ with *Symbiodinium* type on the same coral host. Mol. Ecol. **19:**1978-1990.

17. **Morsy MR, Oswald J, He J, Tang YH, Roossinck MJ.** 2010. Teasing apart a three-way symbiosis: transcriptome analyses of *Curvularia protuberata* in response to viral infection and heat stress. Biochem. Biophys. Res. Commun. **401:**225-230.

18. **Russell JA, Moran NA.** 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proc. Biol. Sci. **273:**603-610.

19. **Oliver KM, Russell JA, Moran NA, Hunter MS.** 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. U.S.A. **100:**1803-1807.

20. **Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB.** 1981. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. Science **213:**340-342.

21. **Weyl EG, Frederickson ME, Yu DW, Pierce NE.** 2010. Economic contract theory tests models of mutualism. Proc. Natl. Acad. Sci. U.S.A. **107:**15712-15716.

22. **Rohwer F, Seguritan V, Azam F, Knowlton N.** 2002. Diversity and distribution of coral-associated bacteria. Mar. Ecol. Prog. Ser. **243:**1-10.

23. **Bosch TCG, McFall-Ngai MJ.** 2011. Metaorganisms as the new frontier. Zoology **114:**185-190.

24. **Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I.** 2007. The role of microorganisms in coral health, disease and evolution. Nat. Rev. Microbiol. **5:**355-362.

25. **Zilber-Rosenberg I, Rosenberg E.** 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol. Rev. **32:**723-735.

26. **Rosenberg E, Zilber-Rosenberg I.** 2011. Symbiosis and development: the hologenome concept. Birth Defects Res C Embryo Today **93:**56-66.

27. **Nyholm SV, Stewart JJ, Ruby EG, McFall-Ngai MJ.** 2009. Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. Environ. Microbiol. **11:**483-493.

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

29. **Olszak T, An DD, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, Blumberg RS.** 2012. Microbial exposure during early life has persistent effects on natural killer T cell function. Science **336:**489-493.

30. **Detournay O, Schnitzler CE, Poole A, Weis VM.** 2012. Regulation of cnidarian-dinoflagellate mutualisms: Evidence that activation of a host TGF beta innate immune pathway promotes tolerance of the symbiont. Dev. Comp. Immunol. **38:**525-537.

31. **Kikuchi Y, Hayatsu M, Hosokawa T, Nagayama A, Tago K, Fukatsu T.** 2012. Symbiont-mediated insecticide resistance. Proc. Natl. Acad. Sci. U.S.A. **109;**8618-22. doi: 10.1073/pnas.1200231109.

32. **Douglas AE.** 1998. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. Annu. Rev. Entomol. **43:**17-37.

33. **Miura T, Braendle C, Shingleton A, Sisk G, Kambhampati S, Stern DL.** 2003. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrthosiphon pisum* (Hemiptera : Aphidoidea). J. Exp. Zool. Part B **295B:**59-81.

34. **Dehority BA.** 2003. Rumen microbiology. Nottingham University Press, Nottingham, UK.

35. **Nyholm SV, Mcfall-Ngai MJ.** 2004. The winnowing: Establishing the squid-*Vibrio* symbiosis. Nat. Rev. Microbiol. **2:**632-642.

36. **Dubilier N, Bergin C, Lott C.** 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat. Rev. Microbiol. **6:**725-740.

37. **Benson DR, Silvester WB.** 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev. **57:**293-319.

38. **Kikuchi Y, Hosokawa T, Fukatsu T.** 2007. Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. Appl. Environ. Microbiol. **73:**4308-4316.

39. **Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL, Gartman A, Luther GW, Seewald JS, Fisher CR, Girguis PR.** 2012. Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. Proc. Natl. Acad. Sci. U.S.A. **109:**E3241-E3250.

40. **Luttge U.** 2012. Modularity and emergence: biology's challenge in understanding life. Plant Biol. (Stuttg) **14:**865-871.

41. **The NIH HMP Working Group, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M.** 2009. The NIH Human Microbiome Project. Genome Res. **19:**2317-2323.

42. **Ibrahim M, Anishetty S.** 2012. A *meta-metabolome* network of carbohydrate metabolism: Interactions between gut microbiota and host. Biochem. Biophys. Res. Commun. **428:**278-284.

43. **Grice EA, Segre JA.** 2011. The skin microbiome. Nat. Rev. Microbiol. **9:**244-253.

44. **Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R.** 2012. Diversity, stability and resilience of the human gut microbiota. Nature **489:**220-230.

45. **Douglas AE.** 2011. Lessons from studying insect symbioses. Cell Host Microbe **10:**359-367.

46. **McCutcheon JP, Moran NA.** 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. **10:**13-26.

47. **Login FH, Balmand S, Vallier A, Vincent-Monegat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A.** 2011. Antimicrobial peptides keep insect endosymbionts under control. Science **334:**362-365.

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

49. **McCutcheon JP, von Dohlen CD.** 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. **21:**1366-1372.

50. **Nikoh N, Hosokawa T, Oshima K, Hattori M, Fukatsu T.** 2011. Reductive evolution of bacterial genome in insect gut environment. Genome Biol. Evol. **3:**702-714.

51. **Rio RVM, Symula RE, Wang J, Lohs C, Wu YN, Snyder AK, Bjornson RD, Oshima K, Biehl BS, Perna NT, Hattori M, Aksoy S.** 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. mBio **3:**e00240-11. doi: 10.1128/mBio.00240-11.

52. **Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, Carninci P, Ishikawa H, Kudo T, Fukatsu T.** 2005. Transcriptome analysis of the aphid

bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. Proc. Natl. Acad. Sci. U.S.A. **102:**5477-5482.

53. **Bermingham J, Rabatel A, Calevro F, Vinuelas J, Febvay G, Charles H, Douglas A, Wilkinson T.** 2009. Impact of host developmental age on the transcriptome of the symbiotic bacterium *Buchnera aphidicola* in the pea aphid (*Acyrthosiphon pisum*). Appl. Environ. Microbiol. **75:**7294-7297.

54. **Burke GR, Moran NA.** 2011. Responses of the pea aphid transcriptome to infection by facultative symbionts. Insect Mol. Biol. **20:**357-365.

55. **Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ.** 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol. Cell Proteomics **10**: M110.007039. doi: 10.1074/mcp.M110.007039.

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

57. **Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H.** 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp* APS. Nature **407:**81-86.

58. **Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SGE.** 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science **296:**2376-2379.

59. **van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A,**

Valencia A, Moran F, Moya A. 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. U.S.A. **100:**581-586.

60. **Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S.** 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. **32:**402-407.

61. **Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA.** 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. **4:**e188. doi:10.1371/journal.pbio.0040188.

62. **McCutcheon JP, Moran NA.** 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. U.S.A. **104:**19392-19397.

63. **McCutcheon JP, McDonald BR, Moran NA.** 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. U.S.A. **106:**15394-15399.

64. **Wernegreen JJ.** 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. **3:**850-861.

65. **Moran NA.** 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U.S.A. **93:**2873-2878.

66. **Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, Aksoy S.** 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. **16:**149-156.

67. **Burke GR, Moran NA.** 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. Genome Biol. Evol. **3:**195-208.

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

69. **Oliver KM, Degnan PH, Burke GR, Moran NA.** 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. Annu. Rev. Entomol. **55:**247-266.

70. **Gosalbes MJ, Lamelas A, Moya A, Latorre A.** 2008. The striking case of tryptophan provision in the cedar aphid *Cinara cedri.* J. Bacteriol. **190:**6026-6029.

71. **Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, Latorre A.** 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: A missing link from facultative to obligate insect endosymbiont. PLoS Genet. **7**: e1002357.

72. **Takiya DM, Tran PL, Dietrich CH, Moran NA.** 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta : Hemiptera : Cicadellidae) and their dual bacterial symbionts. Mol. Ecol. **15:**4175-4191.

73. **McCutcheon JP, Moran NA.** 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol. Evol. **2:**708-718.

74. **von CD, Kohler S, Alsop ST, McManus WR.** 2001. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. Nature **412:**433-436.

75. **Leak S.** 1999. Tsetse biology and ecology. CABI Publishing, Nairobi, Kenya.

76. **FAO.** 1992. *In* Pollock JN (ed), Training manual for tsetse control personnel, vol 1. Food and Agriculture Organization of the United Nations, Rome, Italy.

77. **Clausen PH, Adeyemi I, Bauer B, Breloeer M, Salchow F, Staak C.** 1998. Host preferences of tsetse (Diptera: Glossinidae) based on bloodmeal identifications. Med. Vet. Entomol. **12:**169-180.

78. **Farikou O, Njiokou F, Simo G, Asonganyi T, Cuny G, Geiger A.** 2010. Tsetse fly blood meal modification and trypanosome identification in two sleeping sickness foci in the forest of southern Cameroon. Acta Trop. **116:**81-88.

79. **Weiss BL, Wang JW, Aksoy S.** 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol. **9:**e1000619.

80. **Osei-Poku J, Mbogo CM, Palmer WJ, Jiggins FM.** 2012. Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. Mol. Ecol. **21:**5138-5150.

81. **Broderick NA, Lemaitre B.** 2012. Gut-associated microbes of *Drosophila melanogaster*. Gut Microbes **3:**307-321.

82. **Tobe SS, Langley PA.** 1978. Reproductive physiology of *Glossina*. Annu. Rev. Entomol. **23:**283-307.

83. **Ma WC, Denlinger DL.** 1974. Secretory discharge and microflora of milk gland in tsetse flies. Nature **247:**301-303.

84. **Balmand S, Lohs C, Aksoy S, Heddi A.** 2013. Tissue distribution and transmission routes for the tsetse fly endosymbionts. J. Invertebr. Pathol. **112:**S116-S122. http://dx.doi.org/10.1016/j.jip.2012.04.002.

85. **Aksoy S.** 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. Int. J. Syst. Bacteriol. **45:**848-851.

86. **Dale C, Maudlin I.** 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int. J. Syst. Bacteriol. **49:**267-275.

87. **O'Neill SL, Gooding RH, Aksoy S.** 1993. Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. Med. Vet. Entomol. **7:**377-383.

88. **Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH.** 2008. How many species are infected with *Wolbachia*? - A statistical analysis of current data. FEMS Microbiol. Lett. **281:**215-220.

89. **Werren JH, Baldo L, Clark ME.** 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. **6:**741-751.

90. **Zhou WG, Rousset F, O'Neill S.** 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. Biol. Sci. **265:**509-515.

91. **Cheng Q, Aksoy S.** 1999. Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. Insect Mol. Biol. **8:**125-132.

92. **Geiger A, Fardeau ML, Grebaut P, Vatunga G, Josenando T, Herder S, Cuny G, Truc P, Ollivier B.** 2009. First isolation of *Enterobacter*, *Enterococcus*, and *Acinetobacter* spp. as inhabitants of the tsetse fly (*Glossina palpalis palpalis*) midgut. Infect. Genet. Evol. **9:**1364-1370.

93. **Geiger A, Fardeau ML, Falsen E, Ollivier B, Cuny G.** 2010. *Serratia glossinae* sp. nov., isolated from the midgut of the tsetse fly *Glossina palpalis gambiensis*. Int. J. Syst. Evol. Microbiol. **60:**1261-1265.

94. **Lindh JM, Lehane MJ.** 2011. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie Van Leeuwenhoek **99:**711-720.

95. **Chen XA, Li S, Aksoy S.** 1999. 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.

96. **Weiss BL, Mouchotte R, Rio RVM, Wu YN, Wu Z, Heddi A, Aksoy S.** 2006. Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. Appl. Environ. Microbiol. **72:**7013-7021.

97. **Medecins Sans Frontieres.** 2012. Human African trypanosomiasis (sleeping sickness), p 63-77. *In* Zabriskie P (ed), Fighting Neglect. http://www.msf.es/fighting-neglect/Fighting%20Neglect_ENG.pdf.

98. **Budd LT.** 1999. DFID-funded tsetse and trypanosome research and development since 1980: Economic analysis, vol 2. Natural Resources International Limited, Aylesford, UK.

99. **Aksoy S, Weiss B, Attardo G.** 2008. Paratransgenesis applied for control of tsetse transmitted sleeping sickness. Adv. Exp. Med. Biol. **627:**35-48.

100. **Weiss B, Aksoy S.** 2011. Microbiome influences on insect host vector competence. Trends Parasitol. **27:**514-522.

101. **Frentiu FD, Robinson J, Young PR, McGraw EA, O'Neill SL.** 2010. *Wolbachia*-mediated resistance to Dengue virus infection and death at the cellular level. PloS One **5:**e13398. doi:10.1371/journal.pone.0013398.

102. **Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, Hugo LE, Johnson KN, Kay BH, McGraw**

EA, van den Hurk AF, Ryan PA, O'Neill SL. 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *Plasmodium*. Cell **139:**1268-1278. doi:10.1016/j.cell.2009.11.042.

103. **Bian G, Xu Y, Lu P, Xie Y, Xi Z.** 2010. The endosymbiotic bacterium *Wolbachia* induces resistance to Dengue virus in *Aedes aegypti*. PloS Pathog. **6:**e1000833. doi:10.1371/journal.ppat.1000833.

104. **Kambris Z, Cook PE, Phuc HK, Sinkins SP.** 2009. Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. Science **326:**134-136.

105. **Cook PE, Hugo LE, Iturbe-Ormaetxe I, Williams CR, Chenoweth SF, Ritchie SA, Ryan PA, Kay BH, Blows MW, O'Neill SL.** 2006. The use of transcriptional profiles to predict adult mosquito age under field conditions. Proc. Natl. Acad. Sci. U.S.A. **103:**18060-18065.

106. **Wang S, Ghosh AK, Bongio N, Stebbings KA, Lampe DJ, Jacobs-Lorena M.** 2012. Fighting malaria with engineered symbiotic bacteria from vector mosquitoes. Proc. Natl. Acad. Sci. U.S.A. **109:**12734-12739.

107. **Durvasula RV, Sundaram RK, Kirsch P, Hurwitz I, Crawford CV, Dotson E, Beard CB.** 2008. Genetic transformation of a Corynebacterial symbiont from the Chagas disease vector *Triatoma infestans*. Exp. Parasitol. **119:**94-98.

108. **Snyder AK, McLain C, Rio RVM.** 2012. The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density via exogenous nutrient provisioning. Appl. Environ. Microbiol. **78:**7792-7797.

109. **Weiss BL, Maltz M, Aksoy S.** 2012. Obligate symbionts activate immune system development in the tsetse fly. J. Immunol. **188:**3395-3403.

110. **Nogge G.** 1976. Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. Experientia **32:**995-996.

111. **Pais R, Lohs C, Wu YN, Wang JW, Aksoy S.** 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Appl. Environ. Microbiol. **74:**5965-5974.

112. **Alam U, Medlock J, Brelsfoard C, Pais R, Lohs C, Balmand S, Carnogursky J, Heddi A, Takac P, Galvani A, Aksoy S.** 2011. *Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans*. PLoS Pathog. **7:**e1002415. doi:10.1371/journal.ppat.1002415.

113. **Weiss BL, Wang J, Maltz MA, Wu Y, Aksoy S.** 2013. Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. PLoS Pathog. **9:**e1003318. dio:10.1371/journal.ppat.1003318.

114. **Wang J, Brelsfoard C, Wu Y, Aksoy S.** 2013. Intercommunity effects on microbiome and GpSGHV density regulation in tsetse flies. J. Invertebr. Pathol. **112:**S32-S39. http://dx.doi.org/10.1016/j.jip.2012.03.028.

115. **Wernegreen JJ.** 2012. Strategies of genomic integration within insect-bacterial mutualisms. Biol. Bull. **223:**112-122.

116. **Welburn SC, Maudlin I, Ellis DS.** 1987. *In vitro* cultivation of *Rickettsia*-like-organisms from *Glossina* spp. Ann. Trop. Med. Parasitol. **81:**331-335.

117. **Farikou O, Njiokou F, Mbida Mbida JA, Njitchouang GR, Djeunga HN, Asonganyi T, Simarro PP, Cuny G, Geiger A.** 2010. 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.

118. **Belda E, Moya A, Bentley S, Silva FJ.** 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics **11:**449. doi:10.1186/1471-2164-11-449.

119. **Weiss BL, Wu YN, Schwank JJ, Tolwinski NS, Aksoy S.** 2008. An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. Proc. Natl. Acad. Sci. U.S.A. **105:**15088-15093.

120. **Maltz MA, Weiss BL, O'Neill M, Wu Y, Aksoy S.** 2012. OmpA-mediated biofilm formation is essential for the commensal bacterium *Sodalis glossinidius* to colonize the tsetse fly gut. Appl. Environ. Microbiol. **78:**7760-7768.

121. **Novakova E, Hypsa V.** 2007. A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiol. Lett. **269:**131-135.

122. **Chrudimsky T, Husnik F, Novakova E, Hypsa V.** 2012. *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. PLoS One **7:**e40354. doi:10.1371/journal.pone.0040354.

123. **Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T.** 2010. Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. Appl. Environ. Microbiol. **76:**3486-3494.

124. **Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T.** 2011. Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera: Scutelleridae). Zool. Sci. **28:**169-174.

125. **Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstadter J, Hurst GD.** 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol. **6:**27. doi:10.1186/1741-7007-6-27.

126. **Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, Sasaki-Fukatsu K, Yoshizawa K, Dale C, Clayton DH.** 2007. Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. Appl. Environ. Microbiol. **73:**6660-6668.

127. **Grünwald S, Pilhofer M, Holl W.** 2010. Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles [Coleoptera: Cerambycidae]. Syst. Appl. Microbiol. **33:**25-34. doi:10.1016/j.syapm.2009.10.002.

128. **Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P.** 1998. Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G + C content of an endocytobiotic DNA. J. Mol. Evol. **47:**52-61.

129. **Toju H, Hosokawa T, Koga R, Nikoh N, Meng XY, Kimura N, Fukatsu T.** 2010. "*Candidatus* Curculioniphilus buchneri," a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. Appl. Environ. Microbiol. **76:**275-282.

130. **Toju H, Tanabe AS, Notsu Y, Sota T, Fukatsu T.** 2013. Diversification of endosymbiosis: replacements, co-speciation and promiscuity of bacteriocyte symbionts in weevils. ISME J. doi:10.1038/ismej.2013.27.

131. **Verbarg S, Fruhling A, Cousin S, Brambilla E, Gronow S, Lunsdorf H, Stackebrandt E.** 2008. *Biostraticola tofi* gen. nov., spec. nov., a novel member of the family *Enterobacteriaceae*. Curr. Microbiol. **56:**603-608.

132. **Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C.** 2012. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. PLoS Genet. **8:**e1002990. doi:10.1371/journal.pgen.1002990.

133. **Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, O'Neill SL, Aksoy S.** 2000. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. Med. Vet. Entomol. **14:**44-50.

134. **Doudoumis V, Tsiamis G, Wamwiri F, Brelsfoard C, Alam U, Aksoy E, Dalaperas S, Abd-Alla A, Ouma J, Takac P, Aksoy S, Bourtzis K.** 2012. Detection and characterization of *Wolbachia* infections in laboratory and natural populations of different species of tsetse flies (genus *Glossina*). BMC Microbiol. **12 Suppl 1:**S3. http://www.biomedcentral.com/1471-2180/12/S1/S3.

135. **Abd-Alla AMM, Parker AG, Vreysen MJB, Bergoin M.** 2011. Tsetse salivary gland hypertrophy virus: hope or hindrance for tsetse control? PLoS Negl. Trop. Dis. **5:** e1220. doi:10.1371/journal.pntd.0001220.

136. **Jaenson TG.** 1978. Virus-like rods associated with salivary gland hyperplasia in tsetse, *Glossina pallidipes*. Trans. R. Soc. Trop. Med. Hyg. **72:**234-238.

137. **Jura WG, Odhiambo TR, Otieno LH, Tabu NO.** 1988. Gonadal lesions in virus-infected male and female tsetse, *Glossina pallidipes* (Diptera: Glossinidae). J. Invertebr. Pathol. **52:**1-8.

138. **Sang RC, Jura W, Otieno LH, Ogaja P.** 1996. Ultrastructural changes in the milk gland of tsetse *Glossina morsitans centralis* (Diptera; Glossinidae) Female Infected by a DNA Virus. J. Invertebr. Pathol. **68:**253-259.

139. **Sang RC, Jura WG, Otieno LH, Mwangi RW, Ogaja P.** 1999. The effects of a tsetse DNA virus infection on the functions of the male accessory reproductive gland in the host fly *Glossina morsitans centralis* (Diptera; Glossinidae). Curr. Microbiol. **38:**349-354.

140. **Sang RC, Jura WG, Otieno LH, Tukei PM, Mwangi RW.** 1997. Effects of tsetse DNA virus infection on the survival of a host fly, *Glossina morsitans centralis* (Diptera; Glossinidae). J. Invertebr. Pathol. **69:**253-260.

141. **Sang RC, Jura WG, Otieno LH, Mwangi RW.** 1998. The effects of a DNA virus infection on the reproductive potential of female tsetse flies, *Glossina morsitans centralis* and *Glossina morsitans morsitans* (Diptera: Glossinidae). Mem. Inst. Oswaldo Cruz **93:**861-864.

142. **Rio RV, Wu YN, Filardo G, Aksoy S.** 2006. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. Proc. Biol. Sci. **273:**805-814.

143. **Hamidou Soumana I, Berthier D, Tchicaya B, Thevenon S, Njiokou F, Cuny G, Geiger A.** 2013. Population dynamics of *Glossina palpalis gambiensis* symbionts, *Sodalis glossinidius*, and *Wigglesworthia glossinidia*, throughout host-fly development. Infect. Genet. Evol. **13:**41-48.

144. **Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM.** 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc. Biol. Sci. **277:**2389-2397.

145. **Walshe DP, Ooi CP, Lehane MJ, Haines LR.** 2009. The enemy within: interactions between tsetse, trypanosomes and symbionts. Adv. In Insect Phys. **37:**119-175. doi:10.1016/S0065-2806(09)37003-4.

146. **Harley JM.** 1971. Comparison of the susceptibility of infection with *Trypanosoma rhodesiense* of *Glossina pallidipes*, *G. morsitans*, *G. fuscipes* and *G. brevipalpis*. Ann. Trop. Med. Parasitol. **65:**185-189.

147. **Moloo SK, Kutuza SB.** 1988. Comparative study on the susceptibility of different *Glossina* species to *Trypanosoma brucei brucei* infection. Trop. Med. Parasitol. **39:**211-213.

148. **Moloo SK, Kabata JM, Sabwa CL.** 1994. A study on the maturation of procyclic *Trypanosoma brucei brucei* in *Glossina morsitans centralis* and *G. brevipalpis*. Med. Vet. Entomol. **8:**369-374.

149. **Moloo SK, Okumu IO, Kuria NM.** 1998. Comparative susceptibility of *Glossina longipennis* and *G. brevipalpis* to pathogenic species of *Trypanosoma*. Med. Vet. Entomol. **12:**211-214.

150. **Moloo SK, Orinda GO, Sabwa CL, Minja SH, Masake RA.** 1999. Study on the sequential tsetse-transmitted *Trypanosoma congolense*, *T. brucei brucei* and *T. vivax* infections to African buffalo, eland, waterbuck, N'Dama and Boran cattle. Vet. Parasitol. **80:**197-213.

151. **Distelmans W, D'Haeseleer F, Kaufman L, Rousseeuw P.** 1982. The susceptibility of *Glossina palpalis palpalis* at different ages to infection with *Trypanosoma congolense*. Ann. Soc. belge Med. Trop. **62:**41-47.

152. **Dale C, Welburn SC, Maudlin I, Milligan PJ.** 1995. The kinetics of maturation of trypanosome infections in tsetse. Parasitology **111:**187-191.

153. **Peacock L, Ferris V, Bailey M, Gibson W.** 2012. The influence of sex and fly species on the development of trypanosomes in tsetse flies. PLoS Negl. Trop. Dis. **6:**e1515. doi:10.1371/journal.pntd.0001515.

154. **Ellis DS, Evans DA.** 1977. Passage of Trypanosoma brucei rhodesiense through the peritrophic membrane of Glossina morsitans morsitans. Nature **267:**834-835.

155. **Gibson W, Bailey M.** 2003. The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. Kinetoplastid Biol. Dis. **2:**1.

156. **Gibson W, Peacock L, Ferris V, Williams K, Bailey M.** 2008. The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. Parasit. Vectors **1:**4. doi:10.1186/1756-3305-1-4.

157. **Hao Z, Kasumba I, Lehane MJ, Gibson WC, Kwon J, Aksoy S.** 2001. Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc. Natl. Acad. Sci. U.S.A. **98:**12648-12653.

158. **Hao Z, Kasumba I, Aksoy S.** 2003. Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). Insect Biochem. Mol. Biol. **33:**1155-1164.

159. **Lehane MJ, Aksoy S, Gibson W, Kerhornou A, Berriman M, Hamilton J, Soares MB, Bonaldo MF, Lehane S, Hall N.** 2003. Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. Genome Biol. **4:**R63.

160. **Hu C, Aksoy S.** 2006. Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*. Mol. Microbiol. **60:**1194-1204.

161. **Nayduch D, Aksoy S.** 2007. Refractoriness in tsetse flies (Diptera: Glossinidae) may be a matter of timing. J. Med. Entomol. **44:**660-665.

162. **Oberle M, Balmer O, Brun R, Roditi I.** 2010. Bottlenecks and the maintenance of minor genotypes during the life cycle of *Trypanosoma brucei*. PLoS Pathog. **6:**e1001023.

163. **Gechere G, Terefe G, Belihu K.** 2012. Impact of tsetse and trypanosomiasis control on cattle herd composition and calf growth and mortality at Arbaminch District (Southern Rift Valley, Ethiopia). Trop. Anim. Health Prod. **44:**1745-1750.

164. **Simarro PP, Jannin J, Cattand P.** 2008. Eliminating human African trypanosomiasis: where do we stand and what comes next? PLoS Med. **5:**e55. doi:10.1371/journal.pmed.0050055.

165. **Welburn SC, Maudlin I, Simarro PP.** 2009. Controlling sleeping sickness - a review. Parasitology **136:**1943-1949.

166. **Van Den Abbeele J, Bourtzis K, Weiss B, Cordon-Rosales C, Miller W, Abd-Alla AM, Parker A.** 2013. Enhancing tsetse fly refractoriness to trypanosome infection - A new IAEA coordinated research project. J. Invertebr. Pathol. **112:**S142-S147. http://dx.doi.org/10.1016/j.jip.2012.07.020.

167. **Maudlin I, Welburn SC, Mehlitz D.** 1990. The relationship between rickettsia-like-organisms and trypanosome infections in natural populations of tsetse in Liberia. Trop. Med. Parasitol. **41:**265-267.

168. **Welburn SC, Arnold K, Maudlin I, Gooday GW.** 1993. Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse-flies. Parasitology **107:**141-145.

169. **Welburn SC, Maudlin I.** 1992. The nature of the teneral state in *Glossina* and its role in the acquisition of trypanosome infection in tsetse. Ann. Trop. Med. Parasitol. **86:**529-536.

170. **Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme W, Hannick L, Aslett MA, Shallom J, Marcello L, Hou LH, Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetter J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, MacLeod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Talton L, Turner CMR, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang SL, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barret BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM.** 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science **309:**416-422.

171. **Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, Chukualim B, Capewell P, MacLeod A, Melville SE, Gibson W, Barry JD, Berriman M, Hertz-Fowler C.** 2010. The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human african trypanosomiasis. PLoS Negl. Trop. Dis. **4:** e658. doi:10.1371/journal.pntd.0000658.

172. **Sienkiewicz N, Jaroslawski S, Wyllie S, Fairlamb AH.** 2008. Chemical and genetic validation of dihydrofolate reductase-thymidylate synthase as a drug target in African trypanosomes. Mol. Microbiol. **69:**520-533.

173. **Wang J, Aksoy S.** 2012. PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. Proc. Natl. Acad. Sci U.S.A. **109:**10552-10557. doi:10.1073/pnas.1116431109.

174. **Alam U, Hyseni C, Symula RE, Brelsfoard C, Wu YN, Kruglov O, Wang JW, Echodu R, Alioni V, Okedi LM, Caccone A, Aksoy S.** 2012. Implications of microfauna-host interactions for trypanosome transmission dynamics in *Glossina fuscipes fuscipes* in Uganda. Appl. Environ. Microbiol. **78:**4627-4637.

175. **Rio RV, Hu Y, Aksoy S.** 2004. Strategies of the home-team: symbioses exploited for vector-borne disease control. Trends Microbiol. **12:**325-336.

176. **Doudoumis V, Alam U, Aksoy E, Abd-Alla AMM, Tsiamis G, Brelsfoard C, Aksoy S, Bourtzis K.** 2013. Tsetse-*Wolbachia* symbiosis: Comes of age and has great potential for pest and disease control. J. Invertebr. Pathol. **112:**S94-S103. doi: 10.1016/j.jip.2012.05.010.

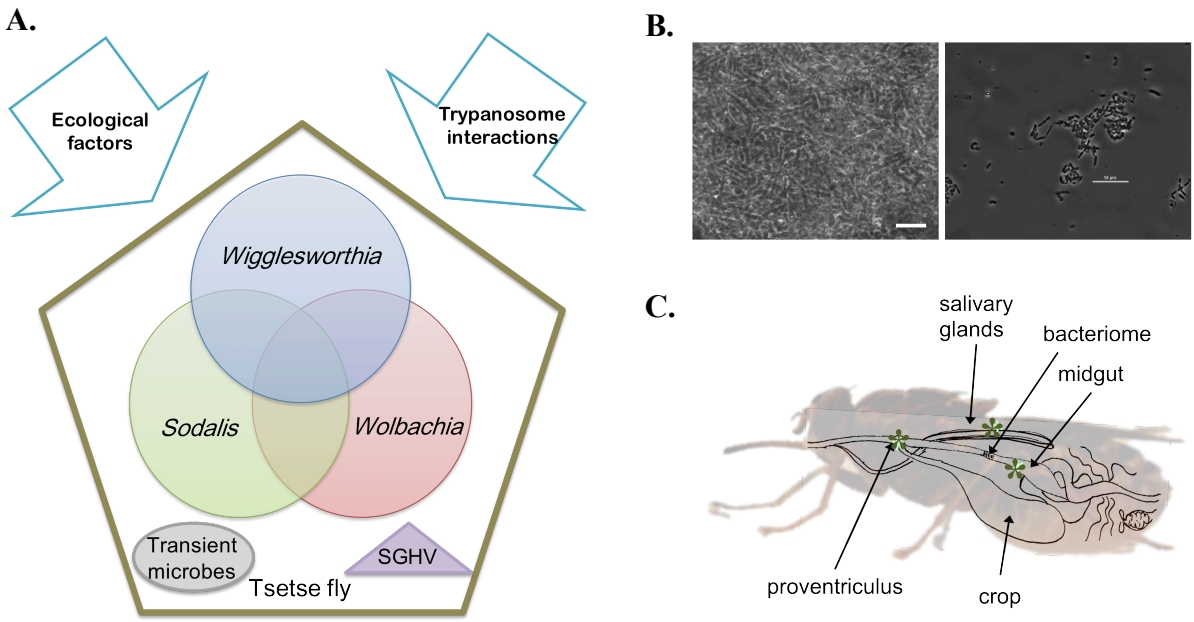


Figure 1-1. The tsetse holobiont. **A.** The holobiont is composed of the tsetse fly and its 3 vertically transmitted bacterial symbionts, *Wigglesworthia, Sodalis and Wolbachia*, and may be influenced by intrinsic factors such as other transient microbes, salivary gland hypertrophy virus (SGHV), and trypanosomes, as well as abiotic factors. **B.** The enteric microbiota (*Wigglesworthia* and *Sodalis*), whose genomes are both annotated, provide a natural model to examine the early evolution of cooperation and adaptations leading towards microbiome co-residency. Phase contrast microscopy images of *Wigglesworthia* cells (left) within a *G. morsitans* bacteriome and *Sodalis* (right) within culture. Scale bars signify 10 µm. **C.** The protozoan parasite *T. brucei subspp*. potentially interacts with microbial symbionts throughout infection. Stages of infection (denoted by *) signify co-localization of microbes that include the midgut, moving anterior towards the proventriculus and culminating in the salivary glands.

Table 1-1. Characterized members of the Enterobacteriaceae *Sodalis*-allied clade exhibiting > 96% 16S rRNA identity.

Members of the Enterobacteriaceae *Sodalis*-allied clade as described in corresponding citation	**Source**	**Insect Order: Family**	**Reference**
Sodalis glossinidius	Tsetse fly, *Glossina spp.*	Diptera: Glossinidae	86
Symbiont	Hippoboscid fly, *Craterina melbae*	Diptera: Hippoboscidae	121
Candidatus Sodalis melophagi	Sheep ked *Melophagus ovinus*	Diptera: Hippoboscidae	122
Sodalis-allied symbiont	Scutellerid stinkbug, *Cantao ocellatus*	Hemiptera: Scutelleridae	123
Sodalis-allied symbiont	Giant jewel stinkbug, *Eucorysses grandis*	Hemiptera: Scutelleridae	124
Symbiont	Long-tailed mealybug, *Pseudococcus longispinus*	Hemiptera: Pseudococcidae	125
Symbiont	Slender pigeon louse, *Columbicola columbae*	Phthiraptera: Philopteridae	126
Symbiont	Longhorn beetle, *Tetropium castaneum*	Coleoptera: Cerambycidae	127
Sitophilus primary symbiont	Grain weevil, *Sitophilus spp.*	Coleoptera: Curculionidae	128
Secondary symbiont	Chestnut weevil, *Curculio sikkimensis*	Coleoptera: Curculionidae	129
Secondary symbiont	Weevil, *Archarius roelofsi*	Coleoptera: Curculionidae	130
Secondary symbiont	Weevil, *Curculio hachijoensis*	Coleoptera: Curculionidae	130
Biostraticola tofi	Environmental: Tufa deposit biofilm isolate	n/a	131
strain HS	Environmental: hand wound	n/a	132

CHAPTER 2: The role of nutrient (Vitamin B1) provisioning in tsetse symbiont homeostasis

CHAPTER 2.1: Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts*

ABSTRACT

Host-associated microbial interactions may involve genome complementation, driving-enhanced communal efficiency and stability. The tsetse fly (Diptera: Glossinidae), the obligate vector of African trypanosomes (*Trypanosoma brucei* subspp.), harbors two enteric Gammaproteobacteria symbionts: *Wigglesworthia glossinidia* and *Sodalis glossinidius*. Host coevolution has streamlined the *Wigglesworthia* genome to complement the exclusively sanguivorous tsetse lifestyle. Comparative genomics reveal that the *Sodalis* genome contains the majority of *Wigglesworthia* genes. This significant genomic overlap calls into question why tsetse maintains the coresidence of both symbionts and, furthermore, how symbiont homeostasis is maintained. One of the few distinctions between the *Wigglesworthia* and *Sodalis* genomes lies in thiamine biosynthesis. While *Wigglesworthia* can synthesize thiamine, *Sodalis* lacks this capability but retains a thiamine ABC transporter (*tbpAthiPQ*) believed to salvage thiamine. This genetic complementation may represent the early convergence of metabolic pathways that may act to retain *Wigglesworthia* and evade species antagonism. We show that thiamine monophosphate, the specific thiamine derivative putatively synthesized by *Wigglesworthia*, impacts *Sodalis*

* Reprinted from: Snyder AK, Deberry JW, Runyen-Janecky L and Rio RVM. 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. R. Soc. B* 277: 2389-2397.

thiamine transporter expression, proliferation and intracellular localization. A greater understanding of tsetse symbiont interactions may generate alternative control strategies for this significant medical and agricultural pest, while also providing insight into the evolution of microbial associations within hosts.

INTRODUCTION

Microbial associations are significant drivers of evolution (20). Since most microbes are localized within a complex consortium, little is known regarding how species interact, and even less is known about mechanisms that prevent species antagonism, which can ultimately compromise the integrity of the biological system. Elucidating these complex microbe – microbe interactions can be enabled through the use of host model systems that harbor naturally simple microbial communities.

The haematophagous tsetse fly (Diptera: Glossinidae) is the sole vector of African trypanosomes (*Trypanosoma brucei* subspp.), the causative agents of the fatal African trypanosomiasis (commonly known as sleeping sickness) in humans and nagana in other animals. In addition to serving as a vector for African trypanosomes, the tsetse fly also harbors two enteric gamma-proteobacterial symbionts: the obligate mutualist *Wigglesworthia glossinidia* (2) and a secondary symbiont, *Sodalis glossinidius* (7). These symbionts are necessary for tsetse's survival, as they are believed to supplement nutrients that the host is incapable of producing or obtaining from its restricted blood diet. The loss of *Wigglesworthia* and *Sodalis* associations results in significant detriment to tsetse, including reduction in reproductive output and shortened lifespan, respectively (26, 25, 8, 28). Tsetse may also harbor *Wolbachia*, which is typically found in reproductive tissues (27) and to date has an unknown functional role. Although examples of fecundity enhancement and greater competitive efficiency have been

described (38, 10, 15), *Wolbachia* associations are generally construed as a form of facultative parasitism within insects (reviewed in 12).

In contrast to the majority of insects, tsetse flies have a unique reproductive strategy known as adenotrophic viviparity (i.e. live birth). Progeny develop through three larval instars *in utero* where they are provided with protein- and lipid-rich nutrients, and also inoculated with *Wigglesworthia* and *Sodalis* via maternal milk gland secretions (4, 19). The maternal transmission of tsetse symbionts is associated with significant population bottlenecks at each generation (30). Consequently, stability between the different microbial symbiont species is especially critical towards maintaining the cohesiveness and evolutionary success of the biological system.

Molecular phylogenetic analysis of the association between *Wigglesworthia* and tsetse supports an ancient establishment, dating back 50 – 80 Myr with a high degree of concordance (5). Extensive host coevolution has streamlined *Wigglesworthia*'s genome to complement the exclusively sanguivorous tsetse lifestyle (1). In comparison to the *Wigglesworthia* symbiosis, molecular phylogenetic analyses date the *Sodalis*–tsetse association to be of recent origin (3, 41). Also supporting its recent transition into symbiosis, *Sodalis* remains one of the few insect symbionts that can still be maintained in culture outside of its host (44).

Large-scale sequencing and annotation has begun to shed light on the functional capabilities of host-associated microbes and their potential roles towards host biology and development (23). The comparative analyses of the annotated *Sodalis* (37) and *Wigglesworthia* (1) genomes enable the identification of complementary pathways of potential metabolic integration. Additionally, during intensive periods of host development, the growth dynamics of *Wigglesworthia* and *Sodalis* mirror one another's, suggestive of intertwined metabolic pathways

(30). Interestingly, the *Sodalis* proteome contains most of the putative *Wigglesworthia* products (i.e. greater than 90% of *Wigglesworthia* coding sequences are orthologues within the *Sodalis* genome). This significant genomic overlap calls into question why tsetse maintains the energetically expensive coresidence of both symbiont species and how symbiont homeostasis is maintained.

The deficiency of B vitamins in blood (13) coupled with the inability of insects to synthesize these essential nutrients (6, 35) suggest their provisioning to tsetse through microbial interactions. One of the few distinctions between the *Sodalis* and *Wigglesworthia* genomes lies in thiamine (vitamin B1) biosynthesis and transport. Although *Wigglesworthia* retains *de novo* thiamine biosynthetic capabilities, *Sodalis* is incapable of its production. To complement its thiamine biosynthetic deficiency, the *Sodalis* genome contains a putative thiamine ABC transport system (TbpAThiPQ), which in other closely related prokaryotes is used to salvage exogenous thiamine (40). We believe that this complementation of genetic inventory between *Wigglesworthia* and *Sodalis* may represent the early convergence of metabolic pathways that may act to ensure the maintenance of the *Wigglesworthia* association while also evading antagonism between the symbiont species.

Here, we examine one aspect of possible interplay between tsetse symbionts: the dependence of *Sodalis* on the provisioning of thiamine by *Wigglesworthia*. We investigate the effect of thiamine and its derivatives towards *Sodalis* proliferation and intracellular localization, a lifestyle feature that is associated with enhanced replication for this microbial symbiont (9). Functional assays characterizing the expression and regulatory patterns of the *Sodalis* thiamine ABC transporter were performed. We present evidence for the necessity of exogenous thiamine towards *Sodalis* fitness, both *in vitro* and within the tsetse fly. The biosynthesis and utilization of

thiamine by *Wigglesworthia* and *Sodalis*, respectively, may be pivotal not only towards the retention of the tsetse–*Wigglesworthia* association, but also to preserve homeostasis of the microbial community within the host. Understanding the metabolic interactions of tsetse symbionts can lead to the identification of novel control strategies towards combating trypanosomiasis prevalence, while also providing insight towards the evolution of microbial associations within hosts.

MATERIALS AND METHODS

(a) *Insects*

Tsetse flies, *Glossina morsitans morsitans*, were maintained at West Virginia University within the Department of Biology insectary at 24 ± 18°C with 50 to 55% relative humidity on a 12/12 h light/dark schedule. Tsetse flies received defibrinated bovine blood (Haemostat, Dixon, CA, USA) every 48 h through an artificial membrane feeding system (22).

(b) *Cell cultures*

Sodalis were isolated from surface-sterilized *G. m. morsitans* pupae and cultured on *Aedes albopictus* C6/36 cells as described previously (7). *Sodalis* were subsequently maintained cell-free *in vitro* at 28°C in Mitsuhashi – Maramorosch (MM) medium (41) supplemented with 5% heat-inactivated fetal bovine serum (FBS). C6/36 cells were maintained in MM medium supplemented with 15% FBS at 28°C.

(c) *Growth assays*

Sodalis growth was compared upon inoculation into six different M9 minimal glucose media types ((33); plus additional supplements as indicated in figure 1a). Log-phase *Sodalis* was diluted to an initial OD_{600} of 0.01. Subsequently, 1 ml of diluted culture was inoculated into 4 ml of each

of the various media types and grown at 28°C without shaking. OD_{600} readings were taken every 24h for 5 days, with three independent trials performed.

(d) *Impact of thiamine monophosphate on* Sodalis *fitness*

Log-phase *Sodalis* was inoculated at an OD_{600} of 0.01 into Media 1 (M9 minimal glucose media +50 μg ml^{-1} Bacto Vitamin Assay Casamino Acids; BD, Franklin Lakes, NJ, USA) with the addition of 0, 50 or 500 μM thiamine monophosphate (TMP; Sigma-Aldrich, St Louis, MO, USA). Every 24 h for 7 days, OD_{600} readings were obtained to measure growth. Three independent trials were performed.

(e) *Analysis of symbiont gene expression* in vitro

To examine the transcription of the *Sodalis* thiamine ABC transporter relative to TMP concentration, we chose to analyze the expression of the *tbpA* gene that encodes the thiamine transporter substrate-binding subunit. RNA was isolated during *in vitro* growth in Media 1 ± 50 μM TMP using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The absence of DNA contamination was verified through PCR using an RNA template lacking a reverse-transcription step. First-strand cDNA synthesis was performed with Superscript III Reverse Transcriptase (Invitrogen), 25 ng random hexamer primers and 200 ng RNA. Real-time quantitative PCR (qPCR) was performed in an iCycler iQ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) using Bio-Rad iQ SYBR Green Supermix, 10 mM of primers (tbpAQTfor and tbpAQTrev; electronic supplementary material, table S1) and 2 μl cDNA template. The amplification settings were an initial 3 min denaturation step at 95.0°C, followed by 40 cycles of 10 s at 95.0°C and 30 s at 54.1°C. Internal standard curves were developed by cloning *tbpA* into the pGEM-T vector (Promega, Madison, WI, USA) using tbpAlongF and tbpAlongR primers (electronic supplementary material, table S1). Quantification of the amplicons relative to the

standard curves was performed using Bio-Rad iCycler $_{I}$Q multi-color real-time PCR optical system software v. 2.0. The respective OD_{600} readings of each time point were used for the normalization of *tbpA* expression. All assays were performed in triplicate and replicates were averaged for each sample.

(f) *Regulation of* Sodalis *thiamine transporter*

The control of the *Sodalis* thiamine ABC transporter by a *thi* box regulatory region was examined using the plasmid-borne *tbpA–gfp* fusions pRJ12, pRJ13 and pRJ14 (figure 3*b*) in wild-type *Escherichia coli* MG1655. To construct the *tbpA– gfp* reporter fusions, primers (UR281 and UR282, UR283 or UR284) were used to amplify promoter DNA from three different *tbpA* regions of the *Sodalis* chromosome (figure 3*b*). The PCR products were digested with *Bam*HI and *Xba*I and cloned into the promoterless *gfp* vector pLR29 (32) to generate pRJ12, pRJ13 and pRJ14, respectively. Overnight cultures of MG1655 containing each respective plasmid were started from freezer stocks inoculated into M9 minimal glucose media and 125 μg ml^{-1} carbenicillin. Cultures were grown overnight at 37°C with shaking. Following overnight growth, each culture was pelleted and resuspended in the original volume of M9 media and carbenicillin. Resuspended cultures (20 μl) were inoculated into 2 ml of M9 media and 125 μg ml^{-1} of carbenicillin + 50 μM TMP. Cultures were grown at 37°C with shaking. At 24 h, 500 μl of each sample was fixed in 2% paraformaldehyde and green fluorescence was quantified using an FACSCaliber (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) fluorescence-activated cell sorter with an excitation at 488 nm to measure single-cell fluorescence. FACSCaliber settings were forward scatter = E01, side scatter = 505 and relative fluorescence between 515 and 545 nm = 798. Three independent trials were performed, with 10000 cells analyzed per sample.

(g) ***Examining the effect of thiamine on*** **Sodalis** ***intracellular replication***

Intracellular localization and proliferation, followed by host cell lysis, is a process associated with *Sodalis* replication both in culture and within the tsetse host (7, 9, 44). To examine the influence of TMP towards intracellular infection and replication by *Sodalis*, C6/36 cells were split into six-well culture plates with MM media + 15% FBS. Log-phase *Sodalis* grown in various media types (Media 1 ± 50 μM TMP or MM media + 5% FBS) were inoculated into a confluent lawn of C6/36 cells at an OD_{600} of 0.01. Prior to inoculation, the supernatant from the wells was replaced with the media used to grow the respective *Sodalis*. To account for any potential effects of the various media types towards C6/36 viability, a replicate of the experiment was performed that lacked *Sodalis* inoculation. At 24 and 168 h post-inoculation, the total well contents (including any adhered C6/36 cells) were aspirated and total DNA isolation was performed using the Holmes–Bonner method (14). The quantification of C6/36 cells was determined through qPCR using the rpL8QTfor and rpL8QTrev primers (electronic supplementary material, table S1), which amplify the *A. albopictus* ribosomal protein (rpL8) gene (GenBank accession no. M99055). The quantification of *Sodalis* density was also determined through qPCR, with corresponding SgexochiQTfor and SgexochiQTrev oligonucleotides (electronic supplementary material, table S1), which amplify the single-copy exochitinase gene (*chi*; GenBank accession no. BSPY11391; (30)). Internal standard curves were developed by cloning amplicons for *A. albopictus* rpl8, using rpl8for and rpl8rev primers (electronic supplementary material, table S1), and *Sodalis chi* was produced with Sgexochifor and Sgexochirev (electronic supplementary material, table S1) into the pGEM-T vector (Promega, Madison, WI, USA), as described previously (30). Quantification of the amplicons relative to the standard curves was performed using SYBR Green I Dye (Bio-Rad) and Bio-Rad

iCycler $_{I}$Q multi-color real-time PCR optical system software v. 2.0. The experiment was performed twice with multiple replicates within each trial.

(h) ***Expression of*** **Sodalis** ***thiamine ABC transporter through tsetse fly development***

Tsetse flies, *G. m. morsitans*, were sacrificed at distinct developmental stages (i.e. late larval, dissected approx. 6 – 9 days in utero; early pupal, less than 48 h post-maternal deposition; late pupal, approx. 28 – 30 days post-maternal deposition; teneral, newly eclosed adults prior to first blood meal; and two-week-old adults). Whole-fly RNA was isolated from single tsetse fly individuals using TRIzol (Invitrogen, Carlsbad, CA, USA) and treated with RNase free – DNaseI (Invitrogen). The absence of DNA contamination was verified using PCR. First-strand cDNA synthesis was performed with 200 ng RNA, a 2 μM primer cocktail of tbpArev and gapDHrev (table S1), and Invitrogen Superscript II Reverse Transcriptase. Second-strand synthesis was performed with the addition of complementary 5' end gene primers (electronic supplementary material, table S1) at 55°C for 35 cycles. The amplification products were analyzed by agarose gel electrophoresis and visualized with Kodak one-dimensional image analysis software. The expression level of endogenous *Sodalis* glyceraldehyde-3- phosphate dehydrogenase (gapDH) within respective time points was used as a loading control.

(i) ***The effect of TMP-supplemented blood meals towards*** **Sodalis** ***thiamine ABC transporter expression within tsetse***

Teneral tsetse were maintained on blood meals supplemented with 50 or 500 μM TMP for two weeks. Whole-fly RNA was isolated from single tsetse individuals using TRIzol, and *tbpA* and *gapDH* reverse-transcriptional analyses were performed as described above.

(j) ***The impact of TMP-supplemented blood meals towards tsetse biology***

Teneral tsetse were maintained on blood meals supplemented with TMP as described previously. Tsetse flies were sacrificed at two weeks of age and DNA isolation performed using the Holmes–Bonner protocol (14). DNA from each experimental sample was analyzed to quantify the density of *Wigglesworthia*, *Sodalis* and *Wolbachia* symbionts as described previously (30).

(k) ***Statistical analysis***

The data were analyzed using JMP 7.0 software (SAS Institute, Cary, NC, USA). A one-way analysis of variance (ANOVA) and Tukey – Kramer *post hoc* pairwise comparison of the mean were performed where appropriate to determine whether symbiont density, thiamine ABC transporter expression or C6/36 density differed between the various treatments. Student's *t*-tests were employed to assess the differences in gfp fluorescence of the plasmid constructs. *F*-tests were applied to assess the homogeneity of variances. The normality of density distributions was determined with a goodness-of-fit test. *Wolbachia* densities were squareroot-transformed to satisfy normality. Significant differences ($p \leq 0.05$) are reported.

RESULTS

(a) ***The role of thiamine in*** **Sodalis** ***fitness***

The growth of *Sodalis* in media supplemented with glucose, vitamins and various thiamine derivatives including thiamine–HCl, thiamine pyrophosphate (TPP) or TMP was observed over 120 h. An increase in the *Sodalis* growth yield was observed with the incremental supplementation of various nutrients, such as glucose and vitamins, to an M9 minimal media base (Media 1). *Sodalis* proliferation increased significantly in cultures supplemented with TMP (figure 1*a*); however, a similar enhancement was not observed upon the addition of other thiamine derivatives (i.e. TPP or thiamine – HCl) to Media 1. These results indicate that *Sodalis*

requires an exogenous thiamine source, preferably in the form of TMP, for optimal growth. Furthermore, *Sodalis* growth is impacted through time, not only by the presence of TMP, but also by different concentrations of this nutrient. A significantly higher mean *Sodalis* density was realized in Media 1 supplemented with 50 μM TMP (ANOVA, $p < 0.0001$; figure 1*b*) in comparison to 500 μM TMP. Moreover, a detrimental growth effect was observed when *Sodalis* was inoculated into Media 1 containing 500 μM TMP. These results demonstrate that *Sodalis* requires exogenous nutrients including TMP for its cultivation outside of the tsetse host.

(b) *Impact of exogenous TMP towards* Sodalis *thiamine ABC transporter expression* in vitro

In free-living bacterial species, exogenous thiamine and its derivatives can be imported into the cell through an ATP-driven thiamine ABC transporter localized to the cell wall (31, 40). At sufficient levels, thiamine and its derivatives can transcriptionally repress further TMP uptake by binding to a riboswitch localized upstream of the thiamine ABC transporter operon known as the *thi* box (46).

To determine whether a similar expression pattern occurs with the *Sodalis* thiamine ABC transporter, we analyzed the expression of *tbpA* in media containing or lacking TMP using qPCR. *Sodalis* grown in media lacking TMP exhibited significantly higher *tbpA* expression than cultures grown in the presence of TMP (ANOVA, $p < 0.001$; figure 1*c*). The significantly higher expression of *tbpA* in cells lacking exogenous TMP through time suggests that these *Sodalis* are attempting to import a vital nutrient for growth via its transporter, and that its functional regulation is intact and similar to that exhibited by free-living bacteria. Interestingly, a similar relationship of decreased tbpA expression through time was observed for *Sodalis* in both media types, suggesting the significance of TMP early in growth.

(c) *Regulation of* Sodalis *thiamine ABC transporter*

The *Sodalis tbpA* promoter has a putative *thi* box (21) at nucleotides 58–97 5' of the transcriptional start site. Thus, based on the high conservation of the *thi* box region upstream of *tbpA* (figure 2a), we hypothesized that the *thi* box still mediates thiamine repression of *Sodalis*'s thiamine ABC transporter. To test this hypothesis, we constructed *tbpA–gfp* fusions (± *thi* box, figure 2b) and examined GFP expression in *E. coli* containing these fusions grown with and without TMP. *Escherichia coli* containing the two *tbpA–gfp* fusions containing the *thi* box (pRJ12 and pRJ13) showed significant reductions (Student's *t*-test, $p<0.05$) – specifically, decreases of 42 and 37%, respectively, in GFP levels when grown in media containing TMP when compared with media lacking TMP (figure 2c). In contrast, there was no statistically significant change in the GFP level with *E. coli* containing the pRJ14 fusion (Student's *t*-test, $p > 0.05$), which lacks the *thi* box, in either media type. These data suggest that the *thi* box remains functionally relevant for TMP regulation of *tpbA* expression by *Sodalis*.

(d) *The effect of TMP towards* Sodalis *intracellular localization and replication*

To determine whether *Sodalis* intracellular replication is compromised when grown in the absence of TMP, a monolayer of *A. albopictus* C6/36 cells was inoculated with *Sodalis* grown in the presence or absence of TMP. This particular cell line has previously been demonstrated to support intracellular localization and subsequent increases in *Sodalis* density (7, 9, 44). To ensure that any changes in the C6/36 density were due solely to *Sodalis* infection and not respective media types, replicate assays lacking *Sodalis* were performed and no effects on C6/36 density were found (data not shown). At 24 h post-inoculation, no significant differences were observed in either C6/36 or *Sodalis* density between the various treatments (data not shown and figure 3b, respectively). As incubation progressed to 168 h, C6/36 density was significantly

lower upon inoculation with *Sodalis* grown in TMP-supplemented media and comparable to when the bacteria are cultured in a rich media base (ANOVA, $p < 0.0001$; figure 3a). Moreover, at the 168 h time point, *Sodalis* density was significantly higher with TMP supplementation than with cells cultured in media lacking this nutrient, supporting an increase in replication rate (ANOVA, $*p < 0.05$; figure 3b). The highest *Sodalis* density was supported with nutrient-rich MM media. This suggests that although TMP is critical for its proliferation, this metabolite is not the sole dietary necessity as additional nutrients further enhance replication (ANOVA, $**p < 0.01$; figure 3b). These results demonstrate that the intracellular infection and subsequent replication of *Sodalis*, typical of its lifestyle within the tsetse fly, is compromised when TMP is lacking.

(e) Sodalis *thiamine ABC transporter expression through tsetse development and upon TMP supplementation of host blood meals*

Semiquantitative reverse-transcriptional analyses of whole tsetse fly RNA reveals that thiamine transport by *Sodalis* is dynamic through host development (figure 4a). Expression levels of *Sodalis tbpA* were highest in the late pupal and teneral adult life stages in both males and females and lowest during the larval and early pupal time points.

Expression levels of *Sodalis tbpA* also demonstrated variability between two-week-old female and male flies, with higher transcriptional activity demonstrated within females (figure 4a,b). We also examined *Sodalis tbpA* expression in tsetse adults maintained on various TMP-supplemented blood meals (figure 4b). The expression of *Sodalis tbpA* decreased in females with greater levels of TMP supplementation in blood meals, while this pattern was not observed within males as augmenting TMP had no effect on transcriptional profiles.

(f) ***The impact of TMP-supplemented blood meals towards symbiont density***

Like other obligate insect mutualists, *Wigglesworthia* is unable to be cultured using *in vitro* methods in the laboratory. Consequently, genetic manipulation is not feasible. We bypassed the inability to mutate *Wigglesworthia* to produce increased levels of TMP by supplementing tsetse blood meals with this vitamin derivative and examining the effects on symbiont density. Because symbionts may contain multiple genomes per cell (16), qPCR was used to determine bacterial genome number by using single copy genes normalized to host single copy genes. In support of previous descriptions (Rio et al. 2006), *Wigglesworthia* abundance was significantly greater within females than males across all treatment groups (ANOVA, $p = 0.01$; figure 4c). Within female tsetse, a higher *Wigglesworthia* density was evident within tsetse maintained on blood only in comparison with TMP-supplemented meals, although statistical significance was lacking (ANOVA, $p = 0.71$). Within males, no differences in *Wigglesworthia* density were found between the various treatments (ANOVA, $p = 0.74$). Interestingly, *Sodalis* was more copious within female tsetse fed blood meals supplemented with 50 μM TMP in comparison with those fed blood only or 500 μM TMP-supplemented blood meals (ANOVA, $p = 0.002$; figure 4d). A similar reduction in *Sodalis* density was observed when tsetse females were fed a higher TMP concentration (i.e. 500 μM TMP; this finding is similar to what we observed with *Sodalis* in culture). No significant differences in *Sodalis* density were observed among the male treatment groups (ANOVA, $p = 0.96$). In addition, *Wolbachia* density did not significantly differ (ANOVA, $p = 0.6$; data not shown) between the various treatment groups within each sex.

DISCUSSION

The significance of microbial interactions within hosts is gaining steadfast recognition (11). Recent studies have demonstrated that symbionts of ancient origin are associated with

genomic complementation, enabling microbial species to reach a synergistic equilibrium that cultivates a highly complex interdependence (47). In contrast to insect associations where symbionts are of ancient origins (24, 36), the tsetse enteric partners have vastly different acquisition times (3, 5), providing a unique opportunity for insight into the adaptation processes associated with early coresidence of microbes within a symbiotic system.

Despite a severely reduced genome (1), *Wigglesworthia* significantly impacts several aspects of tsetse fly biology including reproduction, blood-meal digestion, temperature sensitivity, immunological processing and vector competence (28, 39). Although *Sodalis* has a relatively large (4.2 Mb) chromosome, a significant degree of genomic decay is apparent, mostly represented in the plethora of pseudogenes. This abundance of pseudogenes results in a diminished coding capacity of only 51%, making the *Sodalis* genome one of the least coding bacterial genomes known to date (37). The majority of pseudogenes are homologues of proteins that have functions related to immunological defense or transport and metabolism of carbohydrates and inorganic ions in free-living bacteria. These functions are probably no longer necessary, given the fidelity of vertical transmission through successive tsetse generations (30).

One of the few distinctions between the *Wigglesworthia* and *Sodalis* genomes lies in thiamine biosynthesis. While *Wigglesworthia* is capable of synthesizing thiamine (figure S1), *Sodalis* lacks this capability. While the genes necessary for thiamine biosynthesis have clearly been eroded within the *Sodalis* genome (37), this biosynthetic inability appears to be circumvented through the retention of genes that encode a thiamine ABC transporter (*tbpAthiPQ*). Other *Sodalis* genome tailoring events have occurred following its transition to a host-associated lifestyle. Such events include the alteration of immunogenic components of its cell membrane—notably a truncated lipopolysaccharide, an absent O antigen and modified outer

membrane protein A (i.e. *ompA*)—which are believed to protect against a systemic host immune response and enable tsetse establishment (42). Additionally, extensive genome divergence between *Sodalis* and closely related *Sitophilus oryzae* primary endosymbiont appears tailored towards acquiring metabolites absent from the restricted diets of their specific (29). It is tempting to postulate that the evolutionary pressures, resulting in the maintenance of the *Sodalis* thiamine ABC transporter over thiamine biosynthesis capability, may be indicative of selection at the host (45) rather than the individual symbiont level, acting to promote microbial homeostasis and ultimately tsetse fitness. Recognizing mechanisms that drive homeostasis between microbial species provides a basis of understanding fundamental molecular processes associated with the selection, regulation and evolution of symbiotic communities.

Many vitamins must be obtained either through diet or microbial interactions. Thiamine, an important cofactor in carbohydrate and amino acid metabolism, is essential for cellular physiology and growth (34). Within various insect groups, thiamine deficiency results in the degeneration of the fat body, stunted larval growth and reduced fertility (6, 35). The exclusive blood diet of tsetse, lacking in B-complex vitamins (particularly thiamine; (13)), coupled with the inability of *Sodalis*, *Wolbachia* and tsetse to synthesize thiamine, supports the provisioning of this essential cofactor exclusively by *Wigglesworthia*. With thiamine biosynthesis being a unique *Wigglesworthia* role, provisioning of this vitamin may be essential for both preventing antagonism between tsetse's microbial symbionts and ensuring the maintenance of this obligate mutualist through time.

We demonstrate that *Sodalis* proliferation, both extra- and intracellular, is nutrient-limited, specifically by TMP. In essence, *Sodalis* population dynamics may be regulated not only by presence or absence of TMP but specifically by varying concentrations of this vitamin

supplied by *Wigglesworthia*. Nutritional interactions, such as the metabolic interplay of thiamine biosynthesis and transport between the tsetse symbionts, may act to stabilize bacterial cohabitation within a host. The expression of the *Sodalis* thiamine ABC transporter, regulated by TMP through a functionally conserved *thi* box, appears to be reflective of host nutritional status. We observed higher expression of the *Sodalis* thiamine ABC transporter in tsetse's late pupal and teneral life stages. These particular developmental stages, demarcated by only 48 h, culminate a long quiescent developmental period consisting of approximately 30 days in the soil during which nutrient supplies have been vastly reduced (18).

Tsetse fly fitness has been shown to influence the susceptibility towards trypanosome infection. Specifically, starvation periods greatly increase the probability of parasite establishment within tsetse (17), with the teneral stage being of highest vector competence (43). The decrease in *Sodalis* thiamine ABC transporter expression in two-week-old adults probably reflects an increase in the *Wigglesworthia* population (30) and, correspondingly, the ability to synthesize TMP at higher levels. Additionally, *Sodalis* transporter expression was both higher and most affected by TMP supplementation of tsetse blood meals within teneral females in comparison with similarly aged males. This phenomenon is probably due to additional female-specific roles such as reproduction and nourishment of intrauterine progeny, both processes that will result in greater demands and competition for available nutrients.

These studies provide insight into a metabolic factor: the provisioning of TMP by the obligate mutualist *Wigglesworthia*, which may aid the maintenance of microbial homeostasis within tsetse. Future studies will focus on identifying *Sodalis* compensatory roles towards tsetse symbiosis and whether these also act to stabilize the symbiont community. Given the critical role of tsetse symbiosis on host physiology and ecology, these associations provide a weak link in

tsetse's biology. A greater understanding of tsetse symbiont interactions may generate alternative biological control methods for use in decreasing the prevalence of African trypanosomiasis.

REFERENCES

1. **Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M. and Aksoy, S.** 2002 Genome sequence of the endocellular obligate symbiont of tsetse, *Wigglesworthia glossinidia*. Nat. Gen. 32, 402–407. (doi:10.1038/ng986)
2. **Aksoy, S.** 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia s*p. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. Int. J. Syst. Bacteriol. **45:**848-851.
3. **Aksoy, S., Chen, X. and Hypsa, V.** 1997. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae). Insect Mol. Biol. **6:**183-190.
4. **Attardo, G. M., Lohs, C., Heddi, A., Alam, U. H., Yildirim, S. and Aksoy, S.** 2008. Analysis of milk gland structure and function in *Glossina morsitans*: Milk protein production, symbiont populations and fecundity. J. Insect Physiol. **54:**1236-1242.
5. **Chen, X. A., Song, L. and Aksoy, S.** 1999. 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.
6. **Craig, R. and Hoskins, W. M.** 1940. Insect biochemistry. Annu. Rev. Biochem. **9:**617-640.
7. **Dale, C. and Maudlin, I.** 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int. J. Syst. Bacteriol. **49:**267-275.
8. **Dale, C. and Welburn, S. C.** 2001. The endosymbionts of tsetse flies- manipulating host-parasite interactions. Int. J. Parasitol. **31:**628-631.

9. **Dale, C., Young, S. A., Haydon, D. T. and Welburn, S. C.** 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc. Natl. Acad. Sci. U.S.A. **98:**1883-1888.

10. **Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M. E. and Bouletreau, M.** 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proc. Natl. Acad. Sci. U.S.A. **98:**6247-6252.

11. **Dethlefsen, L., McFall-Ngai, M. J. and Relman, D. A.** 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature **449:**811-818.

12. **Dobson, S. L.** 2003. *Wolbachia pipientis*: impotent by association, p. 199-209. *In* T. A. Bourtzis K. and Miller (ed.), *Insect symbiosis*. CRC Press, Boca Raton, FL.

13. **Edwards, M. A., M Kaufman, M. L. and Storvick, C. A.** 1957. Microbiologic assay for the thiamine content of blood of various species of animals and man. Am. J. Clin. Nutr. **5:**51-55.

14. **Holmes, D. S. and Bonner, J.** 1973. Preparation, molecular-weight, base composition, and secondary structure of giant nuclear ribonucleic-acid. Biochemistry **12:**2330-2338.

15. **Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X. and Fukatsu, T.** 2010. *Wolbachia* as a bacteriocyte associated nutritional mutualist. Proc. Natl. Acad. Sci. U.S.A. **107:**769-774.

16. **Komaki, K. and Ishikawa, H.** 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem. Mol. Biol. **30:**253-258.

17. **Kubi, C., Van Den Abeele, J., De Deken, R., Marcotty, T., Dorny, P. and Van Den Bossche, P.** 2006. The effect of starvation on the susceptibility of teneral and non-teneral tsetse flies to trypanosome infection. Med. Vet. Entomol. **20:**388-392.

18. **Leak, S.** 1999. *Tsetse biology and ecology, their role in the epidemiology and control of trypanosomes*. CABI Publishing, New York, NY.

19. **Ma, W. C. and Denlinger, D. L.** 1974. Secretory discharge and rnicroflora of milk gland in tsetse flies. Nature **247:**301-303.

20. **Margulis, L. and Fester, R.** 1991. *Symbiosis as a source of evolutionary innovation : speciation and morphogenesis*. MIT Press, Cambridge, MA.

21. **Miranda-Rios, J., Morera, C., Taboada, H., Davalos, A., Encarnacion, S., Mora, J. and Soberon, M.** 1997. Expression of thiamin biosynthetic genes (*thi*COGE) and production of symbiotic terminal oxidase cbb3 in *Rhizobium etli.* J. Bacteriol. **179:**6887-6893.

22. **Moloo, S. K.** 1971. An artificial feeding technique for *Glossina*. Parasitology **63:**507-512.

23. **Moran, N. A., McCutcheon J. P. and Nakabachi, A.** 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. **42:**165-190.

24. **Moran, N. A., Tran P. and Gerardo, N. M.** 2005. Symbiosis and insect diversification: an ancient symbiont of sap feeding insects from the bacterial phylum Bacteroidetes. Appl. Environ. Microbiol. **71:**8802-8810.

25. **Nogge, G.** 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82:**101-104.

26. **Nogge, G.** 1976. Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. Experientia **32:**995-996.

27. **O'Neill, S. L., Gooding R. H. and Aksoy, S.** 1993. Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. Med. Vet. Entomol. **7:**377-383.

28. **Pais, R., Lohs, C., Wu, Y., Wang, J. and Aksoy, S.** 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Appl. Environ. Microbiol. **74:**5965-5974.

29. **Rio, R. V., Lefevre, C., Heddi, A. and Aksoy, S.** 2003. Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. Appl. Environ. Microbiol. **69:**6825-6832.

30. **Rio, R. V., Wu, Y. N., Filardo, G. and Aksoy, S.** 2006. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. Proc. R. Soc. B **273:**805-814.

31. **Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. and Gelfand, M. S.** 2002. Comparative genomics of thiamin biosynthesis in procaryotes. J. Biol. Chem. **277:**48949-48959.

32. **Runyen-Janecky, L. J. and Payne, S. M.** 2002. Identification of chromosomal *Shigella flexneri* genes induced by the eukaryotic intracellular environment. Infect. Immun. **70:**4379-4388.

33. **Sambrook, J. and Russell, D. W.** 2001. *Molecular cloning: a laboratory manual.* 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

34. **Schowen, R. L.** 1998. Thiamin-dependent enzymes. In L. Sinnott (ed) *Comprehensive biological catalysis*, vol. 2. Academic Press, San Diego, CA.

35. **Sweetman, M. D. and Palmer, L. S.** 1928. Insects as test animals in vitamin research. J. Biol. Chem. **77:**33-52.

36. **Takiya, D. M., Tran, P. L., Dietrich, C. H. and Moran, N. A.** 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta : Hemiptera : Cicadellidae) and their dual bacterial symbionts. Mol. Ecol. **15:**4175-4191.

37. **Toh, H., Weiss, B. L., Perkin, S. A., Yamashita, A., Oshima, K., Hattori, M. and Aksoy, S.** 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. **16:**149-156.

38. **Wade, M. J. and Chang, N. W.** 1995. Increased male fertility in *Tribolium confusum* beetles after infection with the intracellular parasite *Wolbachia*. Nature **373:**72-74.

39. **Wang, J., Wu, Y., Yang, G. and Aksoy, S.** 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. Proc. Natl. Acad. Sci. U.S.A. **106:**12133-12138.

40. **Webb, E., Class, K. and Downs, D.** 1998. *thi*BPQ encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. J. Biol Chem. **273:**8946-8950.

41. **Weiss, B. L., Mouchotte, R. M., Rio, R. V. M., Wu, Y., Wu, Z., Heddi, A. and Aksoy, S.** 2006. Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. Appl. Environ. Mirobiol. **72:**7013-7021.

42. **Weiss, B. L., Weng, Y., Schwank, J. J., Tolwinski, N. S. and Aksoy, S.** 2008. An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. Proc. Natl. Acad. Sci. U.S.A. **105:**15088-15093.

43. **Welburn, S. C. and Maudlin, I.** 1992. The nature of the teneral state in *Glossina* and its role in the acquisition of trypanosome infection in tsetse. Ann. Trop. Med. Parasitol. **86:**529-536.

44. **Welburn, S. C., Maudlin, I. and Ellis, D. S.** 1987. *In vitro* cultivation of *Rickettsia*-like-organisms from *Glossina* spp. Ann. Trop. Med. Parasitol. **81:**331-335.

45. **Wernegreen, J. J. and Moran, N. A.** 2000. Decay of mutualistic potential in aphid endosymbionts through silencing of biosynthetic loci: *Buchnera* of *Diuraphis*. Proc. R. Soc. B **267:**1423-1431.

46. **Winkler, W., Nahvi, A. and Breaker, R. R.** 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature **419:**952-956.

47. **Wu, D., *et al.*** 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. **4:**1079-1092.

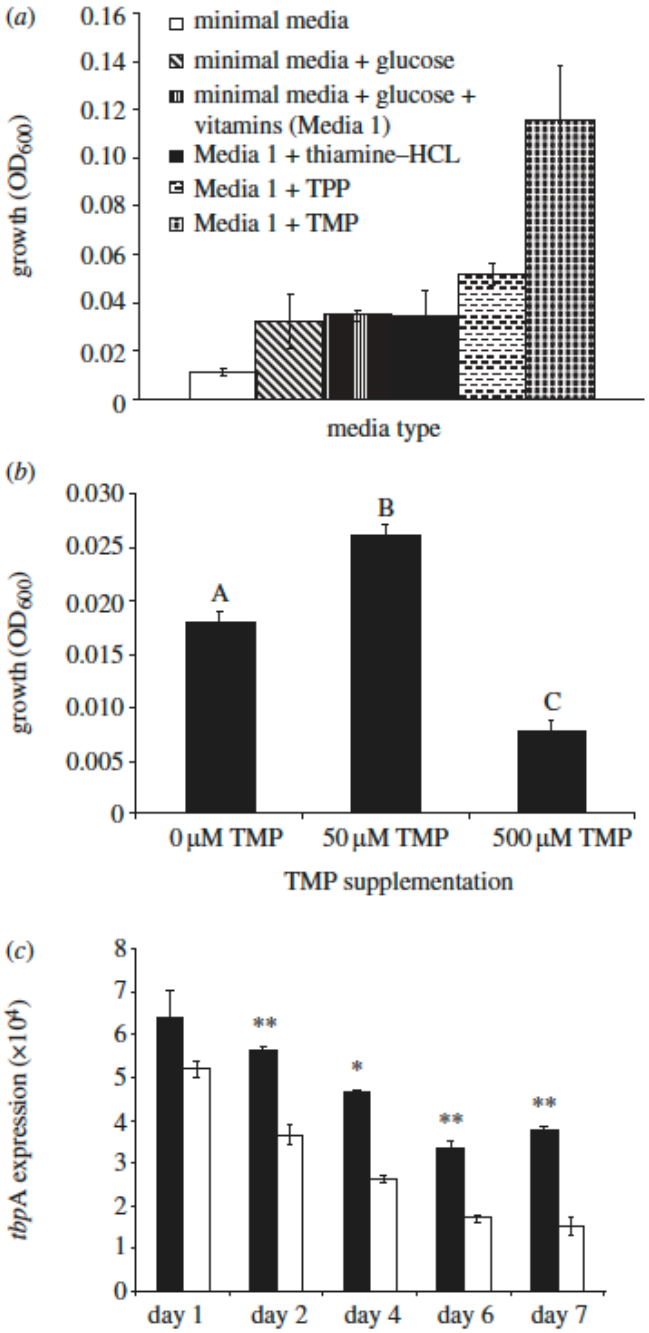

(a)
minimal media
minimal media + glucose
minimal media + glucose + vitamins (Media 1)
Media 1 + thiamine–HCL
Media 1 + TPP
Media 1 + TMP
growth (OD600)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
media type
(b)
growth (OD600)
0.030
0.025
0.020
0.015
0.010
0.005
0
A
B
C
0 μM TMP
50 μM TMP
500 μM TMP
TMP supplementation
(c)
tbpA expression (×10⁴)
8
7
6
5
4
3
2
1
0
**
*
**
**
day 1
day 2
day 4
day 6
day 7

Figure 2.1-1. *Sodalis* growth and thiamine ABC transporter expression *in vitro* within TMP supplemented minimal media. (A) *Sodalis* growth in M9 minimal media ± glucose ± vitamins ± various thiamine derivatives (100 μM) at 120 h post inoculation (vitamins contain a negligible amount of thiamine). (B) Mean *Sodalis* growth through 168 h in Media 1 and 0, 50, or 500 μM TMP. Letters designate treatments that are significantly different from others (ANOVA, $p < 0.0001$). (C) qRT-PCR expression analysis of *tbp*A expression from *Sodalis* grown in Media 1 ± 50 μM TMP. * and ** denote statistically significant differences, ANOVA $p < 0.05$ and $p < 0.01$, respectively, within each time point. Error bars signify ±1 standard error of the mean (s.e.m.). (n ≥ 6 samples at each time point). Black bars, 0 μM TMP; white bars, 50 μM TMP.

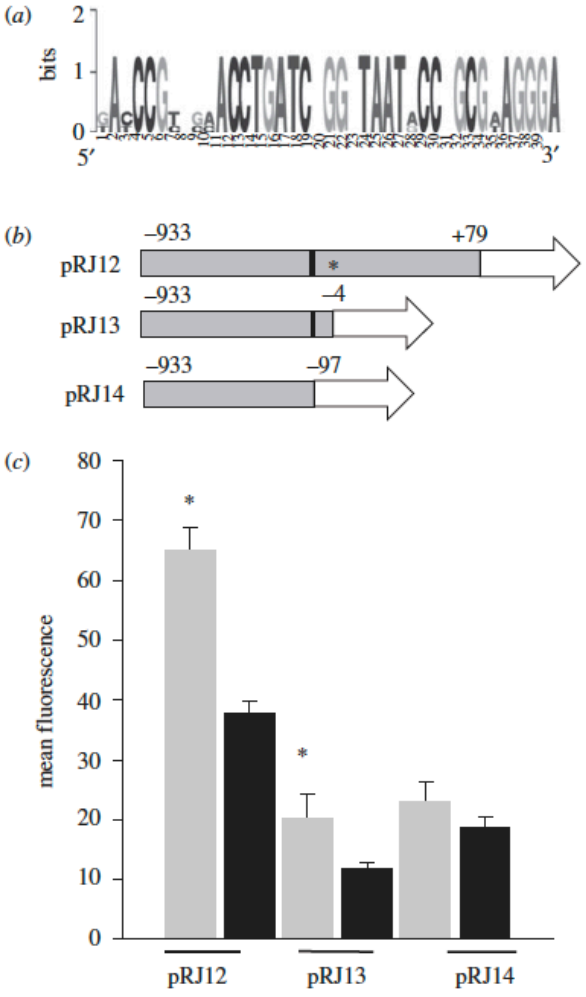

(a)
bits
2
1
0
5′
3′
(b)
–933
+79
pRJ12
*
–933
–4
pRJ13
–933
–97
pRJ14
(c)
mean fluorescence
80
70
60
50
40
30
20
10
0
*
*
pRJ12
pRJ13
pRJ14

Figure 2.1-2. Conservation of *thi* box and regulatory regions of *Sodalis* thiamine ABC transporter. (A) Graphical representation of *thi* box (21) nucleotide sequence alignment of *Sodalis*, *E. coli* and *Salmonella typhimurium*. Image generated through the WEBLOGO website (http://weblogo.berkeley.edu/). (B) *tbpA-gfp* fusions are depicted. The *thi* box and *tbpA* start codon are represented by a black box and an asterisk, respectively. The *gfp* sequences are represented by arrows. (C) *Escherichia coli* MG1655 carrying the *tbp*A-*gfp* fusions were grown for 24 h in Media 1 and carbenicillin in the absence (grey bars) or presence (black bars) of 50 μM TMP and the fluorescence quantified with FACS. Asterisks denote significant differences, Student's *t* –test, $p < 0.05$, within each construct. The data represent the mean fluorescence of at least three independent trials, with 10,000 cells analyzed per sample. Standard deviations are indicated.

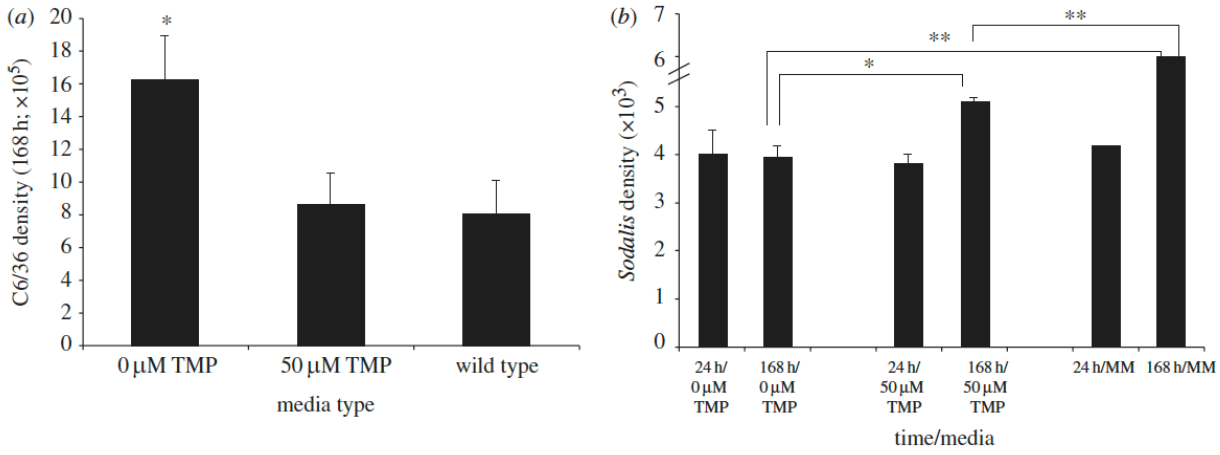


Figure 2.1-3. *Sodalis* intracellular invasion and replication is significantly lower in the absence of TMP. (A) C6/36 density at 168 h post-inoculation with *Sodalis* grown in Media 1 (± 50 µM TMP) or wild type media (MM media + 5% FBS). Mean C6/36 density values are represented and errors bars signify 1 s.e.m. Asterik denotes significant difference (ANOVA, $p < 0.0001$). (B) *Sodalis* density at 24 h and 168 h post inoculation of C6/36 cells. Mean *Sodalis* density values are represented and errors bars signify 1 s.e.m. * and ** denote significant difference, ANOVA $p < 0.05$ and $p < 0.01$, respectively ($n \geq 6$ samples per treatment).

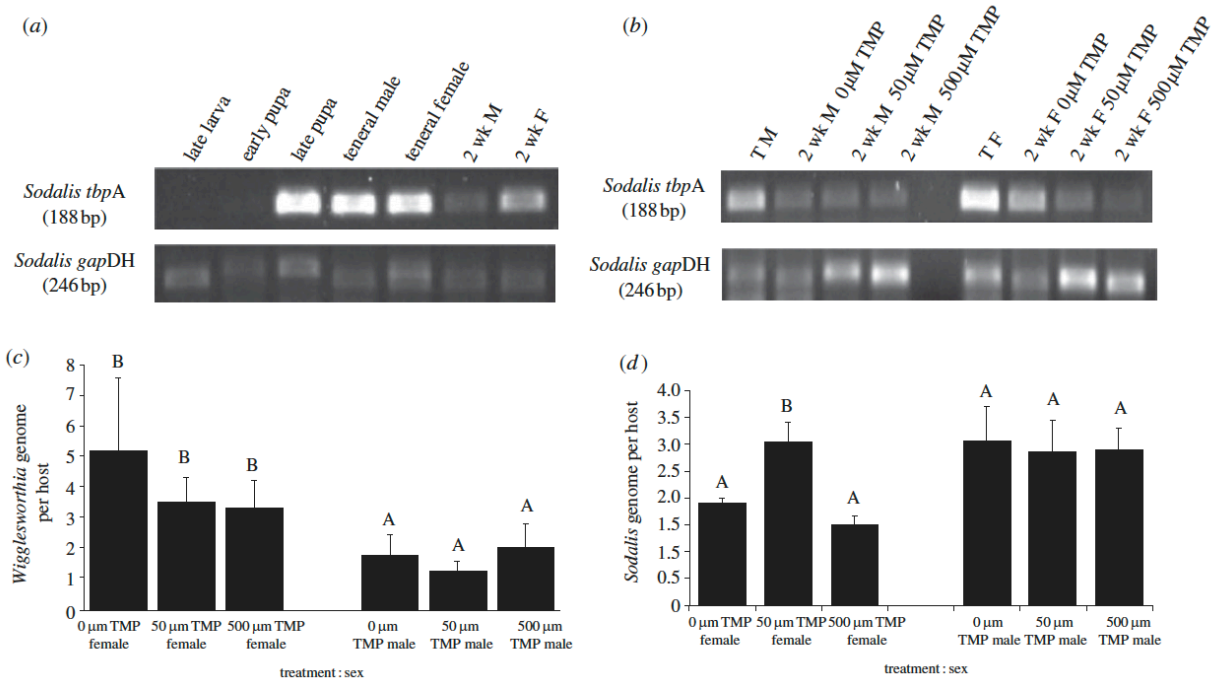


Figure 2.1-4. *Sodalis* thiamine ABC transporter expression and tsetse symbiont density through host development and with supplementation of blood meal. Semiquantitative RT-PCR analysis of *Sodalis tbpA* expression (A) through host development and (B) following two-week TMP supplementation of blood meals. TM, teneral male; 2 wk M, two-week-old male; TF, teneral female; 2 wk F, two-week-old female. *Sodalis gapDH* expression served as a loading control. (C) *Wigglesworthia* and (D) *Sodalis* density were compared in two-week old tsetse fed blood only and TMP supplemented meals. Mean density values are represented and error bars signify 1 s.e.m. Letters depict significant differences (ANOVA, $p < 0.05$) between treatments (n $\geq$ 3 samples at each time point).

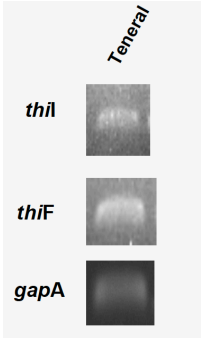


Figure 2.1-S1. *Wigglesworthia* thiamine biosynthesis genes are transcriptionally active. The transcriptional activity of different *Wigglesworthia* thiamine biosynthesis genes was assessed by performing semiquantitative reverse transcription (RT)-PCR. Total RNA was extracted by using Trizol Reagent (Invitrogen) from the bacteriome of a male teneral fly. Negative RT-PCR was used to confirm that all contaminating DNA was removed from the RNA samples following DNAse I (Ambion) treatment. First strand cDNA synthesis was performed with 200 ng RNA, a 2 μm primer cocktail corresponding to the 3' ends of *Wigglesworthia morsitans thiI*, *thiF* and *gapA*, and Invitrogen Superscript II Reverse Transcriptase according to the manufacturer's protocol. Second strand synthesis was performed with the addition of complementary 5' end gene primers at an annealing temperature of 52°C for 30 cycles. The amplification products were analyzed by agarose gel electrophoresis and visualized with Kodak 1D image analysis software. The expression level of endogenous *Wigglesworthia morsitans* glyceraldehyde 3-phosphate dehydrogenase (i.e. *gapA*) was used as a loading control. The specific primers used were as follows: *thiI* (F; 5'-CGCTGAAATACCATATTTTCAAGA-3', R; 5'-TCCTTTTTGGTATAAATATATCGCTTG-3'), *thiF* (F; 5'-TAGCATCTGCAGGGATTGGT-3', R; 5'- TTAACGGTTTATTTTCACTTACACA-3') and *gapA* (F; 5'-GCACCTCCACATGACAACAC-3', R; 5'-TTGCATGAATTG CCCATCTA-3').

Table 2.1-S1. Primers used in this study.

Primer	Primer sequence (5'-3')	Amplicon size	Annealing temperature	Source
gapDHfor	GAG AAC GGT CAT CTG GTG GT	246 bp	55.0 °C	1
gapDHrev	CTG GCC ATC ATA GGC TTT GT			1
rpL8for	GAG GCG AAG TTT GCG TTT AG	899 bp	52.0 °C	1
rpL8rev	GCG TAG TTA CCC GAG GTA CG			1
rpL8QTfor	TTC CGT GAC CCT TAC AAG	172 bp	54.8 °C	1
rpL8QTrev	TCT TCT CCT CCA GAT TGC			1
Sgexochifor	ATG AGC AGT CAG TTA ATT CA	1545 bp	50.0 °C	1
Sgexochirev	CTC ACC AGT GAT ATT AAT CC			1
tbpAlongF	CAA ACC CAT CCT GAC GGT AT	865 bp	55.0 °C	1
tbpAlongR	CTG CGG GGT AAA GCT TAA TG			1
tbpAfor	TCG GAC ATG GTG CTA AGC TA	188 bp	55.0 °C	1
tbpArev	AAA GCC GGG GTA AGG ATA AA			1
tbpAQTfor	GGA ATC GGA CAT GGT GCT AAG	110 bp	54.1 °C	1
tbpAQTrev	ACC TGC TGA TAG TGA CCT TCG			1
UR281	GCG CGG ATC CGA AAT TCG ACA GGA AGA AGG G		55.0 °C	1
UR282	CTA GTC TAG ATA AAT TGC ATA AAG CGA TGG G		55.0 °C	1
UR283	CTA GTC TAG AGC GCT CCT AAG GAC AAG GTA G		55.0 °C	1
UR284	CTA GTC TAG ACT TCT CTC AGC TCA CTC CGA C		55.0 °C	1

1 Designed in this study using either Beacon Designer 7.2 software (Premier Biosoft International) or Primer 3 v.0.4.0 for qPCR and conventional PCR, respectively.

CHAPTER 2.2: The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density with exogenous nutrient provisioning*

ABSTRACT

The obligate mutualist *Wigglesworthia morsitans* provisions nutrients to tsetse flies. The symbiont's response to thiamine (B1) supplementation of blood meals, specifically towards the regulation of thiamine biosynthesis and population density, is described. Despite an ancient symbiosis and associated genome tailoring, *Wigglesworthia* responds to nutrient availability, potentially accommodating a decreased need.

INTRODUCTION

The basis of many microbial symbioses, particularly those within a host, involves nutrient provisioning (29). The spatial co-occurrence of host-associated microbes has been demonstrated to involve several levels of intimacy, including the complementation of biosynthetic capabilities (24, 25, 65). An example of such metabolic interdependency is the provisioning of vitamins by one symbiont species, whereby a second produces essential amino acids, as described in insects within the *Auchenorrhyncha* suborder, such as cicadas and sharpshooters (24, 65) that subsist solely on a diet consisting of plant xylem. An even more extensive level of genetic complementation is exemplified through the integration of gene products from different symbiont species within a single metabolic pathway used in either cross-feeding (8) or the production of nutrients (20, 26). For example, within the mealybug, the synthesis of the essential

* Reprinted from: Snyder AK, McLain C, and Rio RVM. 2012. The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density with exogenous nutrient provisioning. *Appl. Environ. Microbiol.* 78: 7792-7797.

amino acid phenylalanine requires metabolic cooperation between two different bacterial symbiont species (26). A similar partnership has also been described for tryptophan production within the cedar aphid (20). Additionally, precursors provided by the host may regulate the biosynthetic capacity of their microbial partner, ultimately influencing the amount of nutrients produced and released (22, 40, 51). Natural selection favoring metabolic integration may be a mechanism by which species avoid antagonism within their host while also optimizing energy efficiency, particularly if other essential products are provided by each partner.

The Black Queen Hypothesis (BQH) (30) highlights requisites for the evolution of cooperation between species despite selection generally favoring selfishness. The BQH states that the foundation of cooperative community evolution may involve the production of a leaky product by one species, inadvertently providing a public resource, followed by relaxed selection on these biosynthetic pathways within the genome of a beneficiary, thus driving interspecies dependency. Although not explicit to the BQH model, an extension of the hypothesis, stating that these dependencies can favor the development of even tighter associations, such as those of obligate co-evolved partnerships exhibiting genomic signatures of cooperation involving the complementary loss of shared diffusible functions, has also been proposed (44).

The tsetse fly (Diptera: Glossinidae) provides an applicable symbiosis model to study microbial interactions and the evolution of mutualism. The tsetse microbial community is composed predominantly of two *Gammaproteobacteria*; an anciently associated obligate mutualist *Wigglesworthia* speices (2) and a more recently established commensal *Sodalis glossinidius* (12). Tsetse flies are of significant medical and socioeconomic importance as the obligate vectors of parasitic African trypanosomes. A unique feature of tsetse biology is their reproductive strategy, referred to as adenotrophic viviparity, where the majority of larval

development occurs *in utero*. Nutritious lipids and proteins are provisioned to the developing larva through female accessory glands known as milk glands in a mechanism reminiscent of mammalian lactation (3, 7). In addition, *Wigglesworthia* and *Sodalis* are vertically transmitted through these milk gland secretions (4, 21). The strict blood-feeding lifestyle of tsetse flies, coupled with the relatively sterile intrauterine development of larvae, is believed to contribute to the retention of a simple community within tsetse flies (50).

Tsetse symbionts have been shown to impact host biology and have undergone genome adaptations resulting from host association (1, 42, 50). Both annotated *Wigglesworthia* genomes (*Wigglesworthia morsitans*, isolated from *Glossina morsitans* (42)) and *Wigglesworthia brevipalpis*, isolated from *Glossina brevipalpis* (1)) have reduced sizes (~0.7 Mb) due to significant population bottlenecks that occur during vertical transmission, contributing to high levels of genetic drift as well as the relaxed selection and purging of loci no longer necessary due to an obligate host association (61) dating back 50 million to 80 million years (9). The loss of the *Wigglesworthia* association results in significant detriment to tsetse flies, notably the reduction in reproductive output within females (32, 38) that can only be partially restored upon provisioning B vitamins to the blood diet (32, 33). *Wigglesworthia* has also been shown to be essential for symbiont-based maturation of host immunity (53, 58-60). Larvae that lack this symbiont are immunologically compromised as adults with low numbers of hemocytes compared to that for age-matched controls (60). Intracellular *Wigglesworthia* strains are housed within bacteriocyte cells localized to the bacteriome organ in the tsetse anterior midgut. While *Sodalis* has a wider tissue tropism (5, 10), it is harbored primarily within the midgut (10). Relative to *Wigglesworthia*, the *Sodalis* genome is greater in size (~4.2 Mb), yet it shows massive decay with a high number of pseudogenes (6, 50). Functional contributions toward tsetse biology by

Sodalis are much less understood.

Wigglesworthia and *Sodalis* exhibit parallel growth patterns through tsetse host development (43), supporting coordinated activities. Comparative genomic analyses reveal that the majority of *Wigglesworthia* genes (~ 90%) have homologs within the *Sodalis* genome. An exception is in thiamine (Vitamin B1) biosynthetic capability, which appears to be exclusive to *W. morsitans* (42, 46) and not possible by *Sodalis* (6). A recent study demonstrates significant transcriptional regulation of the thiamine biosynthesis locus *thiC* by *W. morsitans* through tsetse development (42). Examples of fine-tuned transcriptional regulation, particularly at the single locus level, that suggest functional and adaptive responses are lacking in other obligate symbionts, such as *Buchnera* in aphids (28, 41, 62, 63) and *Blochmannia* in ants (48).

When grown in minimal medium with and without the presence of thiamine derivatives, *Sodalis* proliferation was shown to require this vitamin, specifically in the form of thiamine monophosphate (TMP) (46). TMP, a physiologically active thiamine derivative, is capable of being produced by *W. morsitans*. Moreover, intracellular invasion and multiplication, an essential feature of *Sodalis* persistence within tsetse flies (13), is also impacted by the availability of exogenous TMP (46). To complement a thiamine biosynthesis deficiency, TMP may be imported by *Sodalis* through a concentration-dependent thiamine ABC transporter (*tbpAthiPQ*) (46). Within tsetse flies, the expression of *Sodalis tbpA*, the thiamine binding protein component of the ABC transporter, was inversely correlated with TMP concentrations, similar to homologs of free-living relatives (55). Furthermore, genetic manipulation aimed at the disruption of the *Sodalis tbpA* locus has proven unsuccessful, suggesting a lethal phenotype (R.V.M. Rio, unpublished data). In addition, *tbpA* transcription, exhibits developmental regulation relative to the tsetse lifecycle (46), with the highest expression occurring at the

conclusion of adult metamorphosis when nutrient supplies are low. Metamorphosis, particularly with holometabolous insects, is a metabolically expensive period during development when adult morphological features are generated without the intake of nutrients. Concordantly, the expression of *W. morsitans thiC* was shown to be highest at this stage in host development (42), potentially indicating this symbiont's response to accommodate a low nutrient environment.

Here, we aim to further understand the dynamics of this symbiont nutrient-provisioning role by examining whether *W. morsitans* remains capable of responding to a lower functional necessity despite a drastically reduced genome and ancient host habitation. This paper details the effects of exogenous vitamin administration of the tsetse blood meal towards *W. morsitans* transcriptional regulation and population proliferation, through the use of gene expression analyses, quantitative PCR (qPCR) and cell viability assays.

METHODS, RESULTS AND DISCUSSION

***W. morsitans* alters gene expression with TMP supplementation.** The obligate mutualism of *W. morsitans* currently hinders isolation in pure culture and downstream applications such as genetic manipulation and subsequent host re-colonization. Therefore, to circumvent this barrier, we supplemented tsetse blood meals with TMP, the thiamine derivative putatively synthesized by *W. morsitans* and previously demonstrated to affect *Sodalis* proliferation and insect cell invasion (46). Male and virgin female *Glossina morsitans morsitans* flies were maintained in the insectary at the Department of Biology at West Virginia University on a 12-h-light//12-h-dark schedule at 24 ± 1°C under 50 to 55% relative humidity. Flies were fed defibrinated bovine blood (Hemostat, Dixon CA) supplemented with 0 to 500 μM TMP every 48 h using an artificial membrane feeding system (27). No significant differences in tsetse fly survival, with three

independent trials performed, were observed between treatment and control groups for the duration of the study period (log rank test, $p = 0.43$ for males, $p = 0.69$ for females).

Thiamine biosynthesis involves the condensation of thiazole and pyrimidine moieties (18). To assess the expression of *W. morsitans* loci involved in thiamine biosynthesis in response to exogenous TMP administration, we examined the thiamine biosynthesis gene *thiC*, involved in synthesis of the pyrimidine moiety, and *thiI*, involved in the formation of the thiazole moiety (18). At 4 weeks after the initial blood meal, semi-quantitative reverse transcription (RT)-PCR was performed to assess the expression of *thiC* and *thiI*. Tsetse bacteriomes were dissected from each treatment group, and RNA was isolated following the TRIzol protocol (Invitrogen, Carlsbad, CA), verified free of DNA contamination, and used for first-strand cDNA synthesis with Superscript II reverse transcriptase (Invitrogen) and a 3' end gene primer cocktail of WgthiCrev (5'-TGC AGC TCC AAT TCC TGA AGT-3'), WgthiIrev (5' - TCC TTT TTG GTA TAA ATA TAT CGC TTG – 3') and WggapArev (5'-TTG CAT GAA TTG CCC ATC TA-3'). The cDNA was then used as template for second-strand synthesis using PCR with WgthiCfor (5'-GAG ATG GTT TGA GAC CTG GAT C-3'; T_a [annealing temperature] = 51°C; 45 cycles; amplicon = 272 bp) and WgthiCrev, WgthiIfor (5' - CGC TGA AAT ACC ATA TTT TCA AGA – 3'; T_a = 55°C; 45 cycles, amplicon = 253 bp) and WgthiIrev, and WggapAfor (5'-GCA CCT CCA CAT GAC AAC AC-3'; T_a = 55°C; 45 cycles; amplicon = 216 bp) and WggapArev for primer sets. *W. morsitans gapA* served to both validate RNA integrity and as a loading control.

Semi-quantitative RT-PCR results demonstrate that as supplemented TMP concentration increases, *W. morsitans thiC* expression correspondingly decreases (Fig. 1A), a transcriptional profile similar to what has been described in free-living bacteria (57). The decrease in *thiC*

expression was more prominent in tsetse males, with reduction first observed with 50 μM TMP, in contrast to 500 μM in females, a finding which may be indicative of the greater need for *W. morsitans* TMP provisioning within females, and the insufficiency of lower TMP concentrations to completely fulfill this demand. In support, the removal of *W. morsitans* has previously been shown to result in female sterility, which can be partially restored upon the provisioning of B vitamins (33). Tsetse female reproductive biology is associated with significant energy demands necessary for oogenesis and other aspects of obligate viviparity (i.e. live birth). In contrast to the dynamics associated with *thiC* expression, no changes in *thiI* expression, in the same individual virgin female samples, were observed in response to TMP supplementation (Fig. 1B). A similar lack of transcriptional regulation of *thiI* by exogenous thiamine has also been observed with *Salmonella enterica* serovar Typhimurium (56), possibly due to its bifunctional role. In addition to its role in thiazole synthesis, ThiI is also involved in 4-thiouridine modification of tRNA (31). Although tRNA modification domains have been lost from *Wigglesworthia* ThiI, the transcription of *thiI* may still be conserved in an unregulated manner. In support, the transcription of *W. morsitans thiI* has also been observed in the teneral (i.e. newly emerged) tsetse adult stage (46).

The ThiI protein is composed of three structural motifs, a THUMP domain, an adenylation domain, and a C-terminal rhodanese-like domain (54). It has recently been demonstrated that the C-terminal rhodanese domain of ThiI is sufficient for thiazole synthesis (23). Interestingly, *W. morsitans thiI* is truncated (~260 nucleotides [nt]) and had previously been classified as a pseudogene (42). Alignment of *W. morsitans* (genome coordinates 670466 to 670200) and *W. brevipalpis* (genome coordinates 172250 to 172543) ThiI amino acid sequence with the rhodanese domains from *Escherichia coli* K-12 (NCBI accession no. YP_001729329),

demonstrates 73.3% and 72.8% similarity levels, respectively, indicating the retention of the sole domain required for thiamine biosynthesis. In further support of their functional conservation, both annotated *Wigglesworthia* genomes contain the critical Arg414 and Cys456 residues in their rhodanese domains (39).

Symbiont population density is impacted by exogenous TMP supplementation. The exogenous administration of TMP in tsetse blood meals for 2 weeks has previously been associated with a decrease in *W. morsitans* population density within female hosts, yet the trend was not significant (46). Therefore, we aimed to explore the effects on cell density by maintaining tsetse on control or 500 μM TMP-supplemented blood meals for a greater temporal period to determine if there is a response observed at the highest TMP concentration previously used to examine tsetse symbiont responses (46). Bacteriomes were dissected from 4-week-old adult tsetse flies, and DNA was isolated using the Holmes-Bonner method (17). qPCR was performed to determine *W. morsitans* genome numbers using the single copy *thiC* gene (43), in a CFX96 real-time PCR Detection System (Bio-Rad, Hercules, CA) using SsoFast EvaGreen Supermix (Bio-Rad), 4.0 μM primers (WgthiCfor and WgthiCrev), and 1 μL DNA as template. The quantification of samples relative to standards was analyzed with Bio-Rad CFX Manager software and normalized to the single-copy *G. morsitans* chitinase gene (66) (NCBI accession no. AF337908; using primers GmchiF [5'- GAG ACA ACA ACT AAT TGG CAC TAC-3'] and GmchiR [5'- GCG TTC ATC GTC ATA ACC TAT CC -3']; amplicon = 97 bp, T_a = 50.3°C) to determine *W. morsitans* genome numbers per host cell, as symbiotic organisms have been shown to contain multiple genomes per cell (19, 43). The assay was performed with ≥ 5 samples per treatment group, and triplicates were averaged for each sample. In concordance with previous studies (43, 46), our results demonstrate a significantly higher density of *W. morsitans* in females

than in males (ANOVA, $p = 0.0125$, data not shown). The greater density in females further supports the significant nutrient provisioning role of this symbiont for this sex (16, 32, 33). The density of *W. morsitans* within the bacteriomes of tsetse flies maintained on TMP-supplemented blood meals was significantly lower in males (Student's *t*-test, $p = 0.009$) (Fig. 2) than in an age-matched control group. A similar trend in the reduction of *W. morsitans* density given TMP supplementation was also observed within the bacteriomes of treated females, yet it lacked statistical significance. Lower *W. morsitans* genome numbers, representative of symbiont population density, with TMP administration may indicate the capability to recognize and respond to a decreased functional necessity within the tsetse host, although more probable host modulation cannot be discounted.

To further explore the trend observed of decreased symbiont density observed via qPCR, which quantifies the abundance of nucleic acids but does not represent cell counts or viability, a LIVE/DEAD *Bac*Light bacterial viability assay (Invitrogen, Eugene, OR) was performed on the bacteriomes of tsetse maintained on control or 500 μM TMP-supplemented blood. This assay uses 2 nucleic acid stains: SYTO 9 dye stains all cells and fluoresces green, and propidium iodide (PI) enters cells with compromised membranes (i.e. dead) and emits a red fluorescence. When PI binds to nucleic acids, there is a displacement of SYTO 9 by PI and the consequential quenching of SYTO 9 by fluorescence resonance energy transfer (FRET) (47).

At 4 weeks after the initial blood meal, individual bacteriomes were dissected, placed into 1 mL 0.85% NaCl, and gently homogenized to release *W. morsitans* cells (43). The bacteria were then centrifuged at 4°C for 5 min at 5000 rpm. The majority of the supernatant was removed, leaving 100 μL solution containing *W. morsitans*. A 1:1 combination of 3.34 mM SYTO 9 dye in dimethyl sulfoxide (DMSO) and 20 mM PI in DMSO was mixed thoroughly and

maintained away from light. Subsequently, 0.3 μL of the dye mixture was mixed with the *W. morsitans* solution, and incubated in the dark at 24°C for 15 min. A slide was then prepared with 5 μL stained bacterial suspension and visualized with fluorescence microscopy, and picture capture was performed using an Olympus FluoView FV1000 confocal laser-scanning microscope (Olympus, Tokyo, Japan). The detection channels were set to Alexa Fluor 488 (excitation: 488 nm; emission: 520 nm) to view SYTO 9 fluorescence and to PI (excitation: 543 nm, emission: 619 nm) to capture PI fluorescence. At least 3 bacteriome samples from each treatment group were examined, with ≥ 5 random fields of view recorded for each sample, and three independent assays performed. The live (green) and dead (red) cells were visualized separately and as an overlay of the filters and quantified directly on microphotography (Fig. 3A). The procedure was also repeated with unstained *W. morsitans* and stained 0.85% NaCl lacking *W. morsitans*, both as negative controls, to verify that there was no auto- or background fluorescence, respectively (data not shown). A two-tailed Fisher's exact test was performed using JMP 7.0 software (SAS Institute, Cary NC) to compare the numbers of live and dead cells between treatments within each sex. Statistically significant differences were found in the number of dead *W. morsitans* cells when comparing bacteriome isolates from both treated female (the Fisher exact test; $P < 0.0001$) and male (the Fisher exact test; $P < 0.0001$) tsetse hosts with age-matched controls (Fig. 3B), with TMP-supplemented groups harboring a higher quantity of dead symbionts. The decrease in the ratio of live to dead cells within bacteriomes provides additional biological evidence that *W. morsitans* populations decrease in light of a lower functional demand. Evidence of altered proliferation due to lower functional necessity can be seen in other symbioses to prevent harmful effects towards the host. For example, the green hydra, *Chlorohydra viridissima*, actively expels their endosymbiotic algae during feeding when

an alternative form of nutrition is available (14). In addition, the facultative endosymbiont of aphids, *Hamiltonella defensa,* harbors a toxin-encoding bacteriophage shown to be instrumental in killing developing parasitoid wasp larvae within infected aphids (35) and remains at intermediate frequencies in natural populations due to costly infection (34, 36). But upon the increased prevalence of parasitoid wasps in the host environment, selection favors the spread of these symbionts as a defense mechanism (34).

CONCLUSIONS

These results demonstrate that although *W. morsitans* has been involved in symbiosis with tsetse flies for a historically significant amount of time (9) and has consequently undergone massive genome shrinkage to accommodate this lifestyle, it appears to still be capable of acclimating to changes in nutrient availability. More specifically, *W. morsitans* not only appears to retain thiamine biosynthetic capability, it is also able to respond to exogenous TMP administration by regulating the transcription of the thiamine biosynthetic locus *thiC*, used in production of the pyrimidine moiety, accordingly. Reductions in *W. morsitans* population density was also observed following vitamin administration, possibly due to either symbiont recognition of a decreased need or, more likely, particularly in light of the drastic genome reduction of *W. morsitans,* through host modulation. In support of more probable host-mediated control, components of the tsetse immune system, notably peptidoglycan recognition protein (PGRP-LB), have been shown to control the abundance of the *W. morsitans* symbiont (52, 53). In further support of host influence, modeling and experimental studies of the *Buchnera* and aphid symbiosis suggests that alterations in nutritional phenotypes may be directed by variation in the host's capacity to supply precursors to symbionts rather than differences in symbiont genomic capabilities (15, 22, 51). Unlike many other anciently associated obligate mutualists that are

spatially segregated within bacteriocytes by a host membrane, *W. morsitans* lies free within the cytosol of host cells (2), perhaps enabling a greater ability to sense and respond to metabolic fluctuations within the host. Within many bacteria, thiamine biosynthesis is regulated through a riboswitch, in which the binding of an effector molecule (i.e. thiamine derivatives) causes allosteric control by feedback inhibition (64). Furthermore, within this riboswitch is a conserved *thi*-box nucleotide sequence, located in the 5' untranslated leader region, which is maintained among a taxonomically wide range of organisms (37). Upon the *in silico* examination of these regions (Riboswitch Explorer; http://www.ibt.unam.mx/biocomputo/conserved_motifs.html) within both available *Wigglesworthia* genomes, no identifiable evidence of a riboswitch is apparent. The involvement of more "global regulators" or small RNAs, which remain to be discovered in the *Wigglesworthia* genomes, may be also responsible for the dynamics in transcriptional profile.

The reductive genome evolution of partners, as described in the BQH (30), can be applied to the tsetse fly symbiosis to further understand changes observed leading toward metabolic dependency among partners. Applying this theory, *W. morsitans* provisions thiamine to supplement the host blood diet, of which some is utilized by *Sodalis* for its growth within tsetse flies. Therefore, if TMP is synthesized as a leaky product by *W. morsitans*, there may be relaxed selection on thiamine biosynthesis loci within the *Sodalis* genome. In support of this, a loss of the *Sodalis* genomic components of this biosynthetic pathway has occurred through evolutionary time, consequently resulting in a metabolic dependency on *W. morsitans* for fitness.

The ability of *W. morsitans* to adapt to nutrient availability within its environment may be vital for maintaining homeostasis within tsetse flies, preventing symbiont populations from becoming too large, which could skew the relationship towards pathogenesis. Thiamine is a

crucial cofactor in amino acid and carbohydrate metabolism; therefore, it is necessary for proper cell growth and physiology in all living organisms (45). Within insects, thiamine deficiency results in the degeneration of the fat body, stunted larval growth and reduced fertility (reviewed in 11, 49). Our results support the hypothesis that TMP is a key metabolite in the maintenance of tsetse symbiont homeostasis, as it has been shown to impact *Sodalis* proliferation (46) and also various aspects of *W. morsitans* biology. Moreover, although a massive reduction in *W. morsitans* genomic content has occurred, thiamine production appears to remain intact within its biosynthetic capabilities. Future studies should examine how these modifications are being recognized, especially given the paucity of environmental sensing capabilities with a small to nonexistent repertoire of one- and two-component signal transduction systems within the *Wigglesworthia* genomes.

REFERENCES

1. **Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S.** 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. **32:**402-407.
2. **Aksoy S.** 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. Int. J. Syst. Bacteriol. **45:**848-851.
3. **Attardo GM, Benoit JB, Michalkova V, Yang G, Roller L, Bohova J, Takáč P, Aksoy S.** 2012. Analysis of lipolysis underlying lactation in the tsetse fly, *Glossina morsitans*. Insect Biochem. Mol. Biol. **42:**360-370.
4. **Attardo GM, Lohs C, Heddi A, Alam UH, Yildirim S, Aksoy S.** 2008. Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity. J. Insect Physiol. **54:**1236-1242.
5. **Balmand S, Lohs C, Aksoy S, Heddi A.** 2012. Tissue distribution and transmission routes for the tsetse fly endosymbionts. J. Invertebr. Pathol. in press.
6. **Belda E, Moya A, Bentley S, Silva FJ.** 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics. **11:**449.
7. **Benoit JB, Yang G, Krause TB, Patrick KR, Aksoy S, Attardo GM.** 2011. Lipophorin acts as a shuttle of lipids to the milk gland during tsetse fly pregnancy. J. Insect Physiol. **57:**1553-1561.
8. **Bomar L, Maltz M, Colston S, Graf J.** 2011. Directed culturing of microorganisms using metatranscriptomics. MBio. **2:**e00012-00011.

9. **Chen X, Li S, Aksoy S.** 1999. 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.

10. **Cheng Q, Aksoy S.** 1999. Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. Insect Mol. Biol. **8:**125-132.

11. **Craig R, Hoskins WM.** 1940. Insect Biochemistry. Annu. Rev. Biochem. **9:**617-640.

12. **Dale C, Maudlin I.** 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int. J. Syst. Bacteriol. **49:**267-275.

13. **Dale C, Young SA, Haydon DT, Welburn SC.** 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc. Natl. Acad. Sci. U.S.A. **98:**1883-1888.

14. **Fishman Y, Zlotkin E, Sher D.** 2008. Expulsion of symbiotic algae during feeding by the green hydra--a mechanism for regulating symbiont density? PloS One **3:**e2603.

15. **Gerardo NM, Wilson AC.** 2011. The power of paired genomes. Mol. Ecol. **20:**2038-2040.

16. **Hill P, Saunders DS, Campbell JA.** 1973. The production of "symbiont-free" *Glossina morsitans* and an associated loss of female fertility. Trans. R. Soc. Trop. Med. Hyg. **67:**727-728.

17. **Holmes DS, Bonner J.** 1973. Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. Biochemistry. **12:**2330-2338.

18. **Jurgenson CT, Begley TP, Ealick SE.** 2009. The structural and biochemical foundations of thiamin biosynthesis. Annu. Rev. Biochem. **78:**569-603.

19. **Komaki K, Ishikawa H.** 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem. Mol. Biol. **30:**253-258.

20. **Lamelas AM, Gosalbes J, Manzano-Marin A, Pereto J, Moya A, Latorre A.** 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. PLoS Genet. **7:**e1002357.

21. **Ma W, Denlinger DL.** 1974. Secretory discharge and microflora of milk gland in tsetse flies. Nature. **247:**301-303.

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

23. **Martinez-Gomez NC, Palmer LD, Vivas E, Roach PL, Downs DM.** 2011. The rhodanese domain of ThiI is both necessary and sufficient for synthesis of the thiazole moiety of thiamine in *Salmonella enterica*. J. Bacteriol. **193:**4582-4587.

24. **McCutcheon JP, McDonald BR, Moran NA.** 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. U.S.A. **106:**15394-15399.

25. **McCutcheon JP, Moran NA.** 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. U.S.A. **104:**19392-19397.

26. **McCutcheon JP, von Dohlen CD.** 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. **21:**1366-1372.

27. **Moloo SK.** 1971. An artificial feeding technique for *Glossina*. Parasitology. **63:**507-512.

28. **Moran NA, Dunbar HE, Wilcox JL.** 2005. Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. J. Bacteriol. **187:**4229-4237.

29. **Moran NA, McCutcheon JP, Nakabachi A.** 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. **42:**165-190.

30. **Morris JJ, Lenski RE, Zinser ER.** 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio. **3:**e00036-00012.

31. **Mueller EG, Buck CJ, Palenchar PM, Barnhart LE, Paulson JL.** 1998. Identification of a gene involved in the generation of 4-thiouridine in tRNA. Nucleic Acids Res. **26:**2606-2610.

32. **Nogge G.** 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82:**101-104.

33. **Nogge G.** 1976. Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. Experientia **32:**995-996.

34. **Oliver KM, Campos J, Moran NA, Hunter MS.** 2008. Population dynamics of defensive symbionts in aphids. Proc. Biol. Sci. **275:**293-299.

35. **Oliver KM, Degnan PH, Hunter MS, Moran NA.** 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science. **325:**992-994.

36. **Oliver KM, Russell JA, Moran NA, Hunter MS.** 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. U.S.A. **100:**1803-1807.

37. **Ontiveros-Palacios N, Smith AM, Grundy FJ, Soberon M, Henkin TM, Miranda-Ríos J.** 2008. Molecular basis of gene regulation by the THI-box riboswitch. Mol. Microbiol. **67:**793-803.

38. **Pais R, Lohs C, Wu Y, Wang J, Aksoy S.** 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Appl. Environ. Microbiol. **74:**5965-5974.

39. **Palenchar PM, Buck CJ, Cheng H, Larson TJ, Mueller EG.** 2000. Evidence that ThiI, an enzyme shared between thiamin and 4-thiouridine biosynthesis, may be a sulfurtransferase that proceeds through a persulfide intermediate. J. Biol. Chem. **275:**8283-8286.

40. **Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ.** 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol. Cell Proteomics. **10:**M110.007039.

41. **Reymond N, Calevro F, Viñuelas J, Morin N, Rahbé Y, Febvay G, Laugier C, Douglas A, Fayard JM, Charles H.** 2006. Different levels of transcriptional regulation due to trophic constraints in the reduced genome of *Buchnera aphidicola* APS. Appl. Environ. Microbiol. **72:**7760-7766.

42. **Rio RVM, Symula RE, Wang J, Lohs C, Wu YN, Snyder AK, Bjornson RD, Oshima K, Biehl BS, Perna NT, Hattori M, Aksoy S.** 2012. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. MBio. **3:**e00240-11.

43. **Rio RVM, Wu YN, Filardo G, Aksoy S.** 2006. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. Proc. Biol. Sci. **273:**805-814.

44. **Sachs JL, Hollowell AC.** 2012. The origins of cooperative bacterial communities. MBio. **3:**e00099-12.

45. **Schowen RL.** 1998. Thiamin-dependent enzymes, p 217-266. *In* Sinnott M (ed) Comprehensive biological catalysis, vol 2. Academic Press, San Diego, CA.

46. **Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM.** 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc. Biol. Sci. **277:**2389-2397.

47. **Stocks SM**. 2004. Mechanism and use of the commercially available viability stain, *Bac*Light. Cytometry A. **61:**189-195.

48. **Stoll S, Feldhaar H, Gross R.** 2009. Transcriptional profiling of the endosymbiont *Blochmannia floridanus* during different developmental stages of its holometabolous ant host. Environ. Microbiol. **11:**877-888.

49. **Sweetman MD, Palmer LS.** 1928. Insects as test animals in vitamin research. J. Biol. Chem. **77:**33-52.

50. **Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S.** 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. **16:**149-156.

51. **Vogel KJ, Moran NA.** 2011. Sources of variation in dietary requirements in an obligate nutritional symbiosis. Proc. Biol. Sci. **278:**115-121.

52. **Wang J, Aksoy S.** 2012. PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring. Proc. Natl. Acad. Sci. U.S.A. **109:**10552-10557.

53. **Wang J, Wu Y, Yang G, Aksoy S.** 2009. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. Proc. Natl. Acad. Sci. U.S.A. **106:**12133-12138.

54. **Waterman DG, Ortiz-Lombardia M, Fogg MJ, Koonin EV, Antson AA.** 2006. Crystal structure of *Bacillus anthracis* ThiI, a tRNA-modifying enzyme containing the predicted RNA-binding THUMP domain. J. Molec. Biol. **356:**97-110.

55. **Webb E, Claas K, Downs D.** 1998. *thiBPQ* encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. J. Biol. Chem. **273:**8946-8950.

56. **Webb E, Claas K, Downs DM.** 1997. Characterization of *thiI*, a new gene involved in thiazole biosynthesis in *Salmonella typhimurium*. J. Bacteriol. **179:**4399-4402.

57. **Webb E, Febres F, Downs DM.** 1996. Thiamine pyrophosphate (TPP) negatively regulates transcription of some *thi* genes of *Salmonella typhimurium*. J. Bacteriol. **178:**2533-2538.

58. **Weiss B, Aksoy S.** 2011. Microbiome influences on insect host vector competence. Trends Parasitol. **27:**514-522.

59. **Weiss BL, Maltz M, Aksoy S.** 2012. Obligate symbionts activate immune system development in the tsetse fly. J. Immunol. **188:**3395-3403.

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

61. **Wernegreen JJ.** 2002. Genome evolution in bacterial endosymbionts of insects. Nat. Rev. Genet. **3:**850-861.

62. **Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA.** 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Mol. Microbiol. **48:**1491-1500.

63. **Wilson AC, Dunbar HE, Davis GK, Hunter WB, Stern DL, Moran NA.** 2006. A dual-genome microarray for the pea aphid, *Acyrthosiphon pisum*, and its obligate bacterial symbiont, *Buchnera aphidicola*. BMC Genomics **7:**50.

64. **Winkler W, Nahvi A, Breaker RR.** 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature. **419:**952-956.

65. **Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA.** 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. **4:**e188.

66. **Yan J, Cheng Q, Narashimhan S, Li CB, Aksoy S.** 2002. Cloning and functional expression of a fat body-specific chitinase cDNA from the tsetse fly, *Glossina morsitans morsitans*. Insect Biochem. Mol. Biol. **32:**979-989.

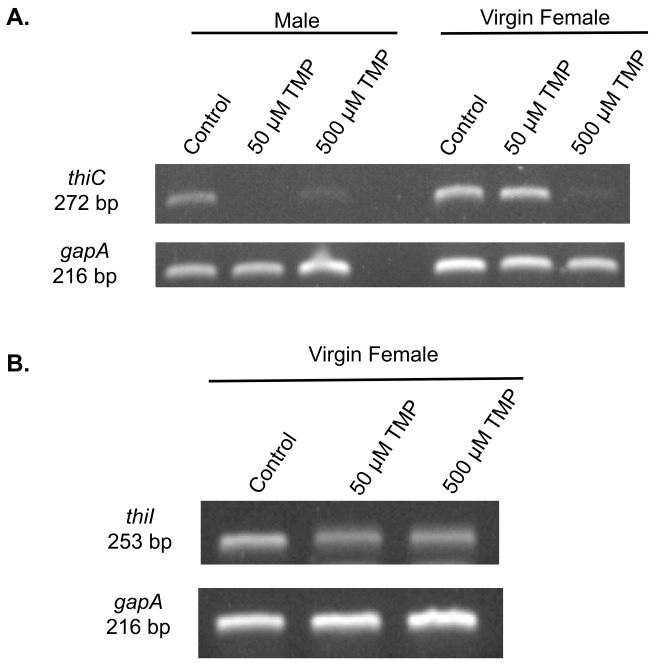


Figure 2.2-1. The expression of *W. morsitans* thiamine biosynthesis genes *thiC* and *thiI* with exogenous TMP supplementation. Semi-quantitative RT-PCR analyses of *W. morsitans thiC* (A) and *thiI* (B) expression in 4-week-old adults maintained with or without TMP supplementation. The constitutively expressed *W. morsitans gapA* was used to verify RNA integrity and as a loading control. At least 3 individual bacteriomes were examined per treatment group.

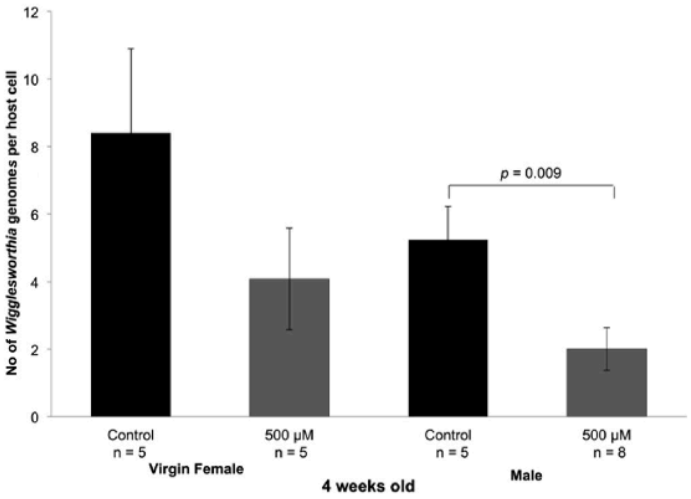


Figure 2.2-2. Dietary thiamine supplementation impacts *W. morsitans* density. *W. morsitans* densities were compared in 4-week-old hosts maintained on control or 500 µM TMP-supplemented blood meals. Mean density values are shown, with error bars signifying 1 standard error of the mean (SEM). Sample sizes are provided below treatment groups.

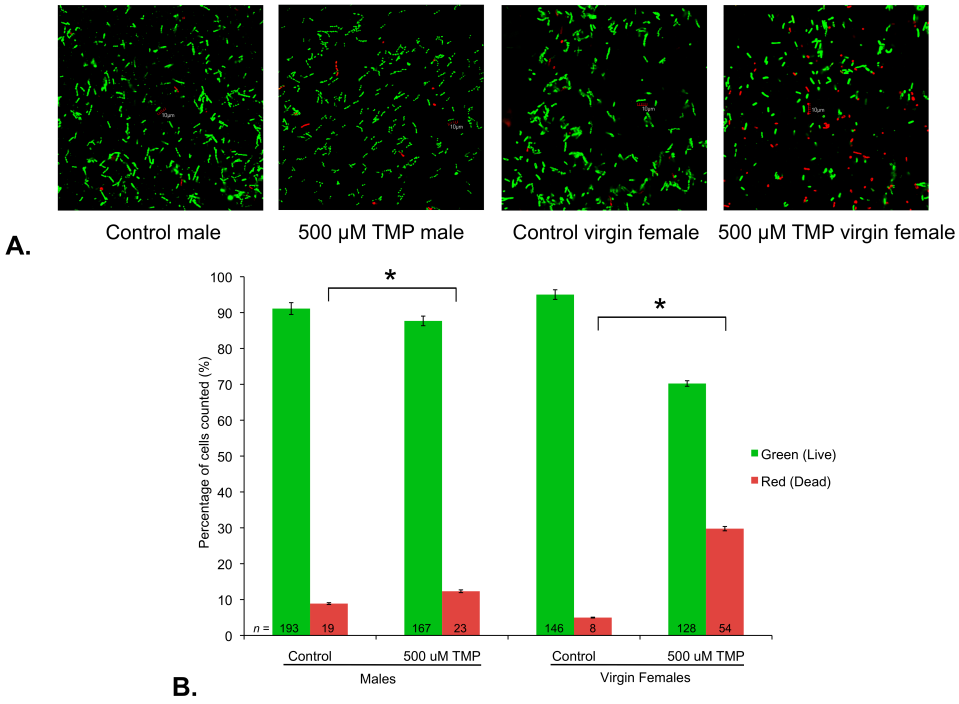


Figure 2.2-3. Viability of *W. morsitans* isolated from the bacteriomes of 4-week-old tsetse flies maintained on TMP-supplemented blood. (A) Representative confocal fluorescence images of *Wigglesworthia* stained with SYTO 9 and PI. (B) Proportions of green (live) and red (dead) cells were quantified in TMP-treated and age- and sex-matched control bacteriomes. Percentages of cells are depicted, with error bars signifying 1 SEM. A total of $\geq$ 3 individuals were analyzed per treatment, with $\geq$ 5 random frames per individual analyzed, and 3 independent assays performed. A two-tailed Fisher exact test was performed to compare the number of live and dead cells between treatments; * $P < 0.0001$. n = average number of cells counted per sample.

CHAPTER 3: The influence of *Wigglesworthia* metabolome distinctions towards tsetse biology

ABSTRACT

Closely related ancient endosymbionts retain genetic distinctions, yet the biological significance of these small pockets of unique loci remains to be described. The tsetse fly (Diptera: Glossinidae), the sole and obligate vector of lethal African trypanosomes, maintains an ancient mutualism with the Gammaproteobacterium *Wigglesworthia* spp.. Tsetse species have undergone extensive concordant evolution with their associated *Wigglesworthia* spp., therefore, the retention of unique symbiont loci may contribute towards host physiological and ecological distinctions. A unique genomic characteristic of *Wigglesworthia morsitans* (Wgm), isolated from *Glossina morsitans*, is the retention of the complete chorismate and downstream folate (Vitamin B9) biosynthesis pathways. Here, we examine the functionality and significance of Wgm chorismate and folate production towards various aspects of host biology to determine whether selection may have preserved these capabilities during co-evolution due to contributory roles in *G. morsitans* biology. These studies highlight the significance of distinctive genomic traits of ancient mutualists towards host biology and phenotypic complexity. The enhanced knowledge of this mutualism may also aid in designing novel avenues for tsetse vector control.

INTRODUCTION

Bacteria adapt to specific environments, including host-associated niches, through the retention or acquisition of functional capabilities during events such as horizontal gene transfer (1, 2) or the establishment of syntrophy (3, 4). Research on ancient obligate associations has demonstrated that microbial symbiont genome evolution can be influenced by microbial

community dynamics (5-10), in addition to host physiology and ecology (11-13). One extreme case has been described within the mealybug where dual ancient symbionts and the host have retained complementary loci that, only when integrated as a symbiotic system (i.e. a holobiont) and not as individual species, are capable of producing specific nutrients (7, 13). Extensive gene purging is characteristic among ancient bacterial symbionts, as they challenge the lower limits of genome size (14, 15). These symbionts exhibit tremendous genomic stasis between strains and species, retaining only those capabilities necessary for the maintenance of the mutualism (Reviewed in (14)). Contrastingly, free-living bacteria adapt to their surroundings by encoding a plethora of strain-specific loci, known as dispensable genes (16), contributing towards their successful viability within alternative environments (17, 18). With ancient bacterial symbionts, much smaller pockets of unique genes have also been observed between species and strains, but to date their functional roles and potential contributions to host phenotypic complexity have not been described.

The tsetse fly (Diptera: Glossina) is a valuable model system to gain deeper insight into the evolution of host associated symbioses due to its low-complexity microbiota. Tsetse can harbor three bacterial symbionts, the obligate mutualist *Wigglesworthia* spp. (19), the commensal *Sodalis glossinidius* (20), and the parasitic *Wolbachia* (21, 22). Field studies report a more complex and diverse adult microbiota (23-25), although the majority of these microbes are believed to be transient in nature. The low complexity of the tsetse microbiota is believed to be maintained by facets of tsetse biology which limit microbial exposure namely; a strict vertebrate blood diet by both sexes, as well as through a unique reproductive strategy, known as adenotrophic viviparity (26). This reproductive mode involves the majority of larval development occurring *in utero*, where the mother provides nourishment, mostly in the form of

amino acids and lipids (27), and vertically transmits the *Wigglesworthia* and *Sodalis* symbionts through highly specialized accessory organs, known as milk glands (28, 29). The significant maternal investment results in a low number of progeny, relative to other Diptera.

In addition to host-associated microbial genome evolution, the enhanced understanding of the tsetse symbiosis also holds medical and agricultural significance, as tsetse flies are the obligate vectors of *Trypanosoma* spp., the causative agents of fatal Human African Trypanosomiasis (HAT) and Nagana, a veterinary wasting disease. While the causes are not fully understood, differences in vector competency (i.e. ability to harbor and transmit trypanosomes) have been observed between sexes and tsetse species (30-36). Immunological responses have been shown to vary between flies that obtain infections and those able to clear a trypanosome challenge (i.e. denoted as refractory) (37-39), yet the contributory roles of symbiont metabolites remain largely understudied.

The tsetse-*Wigglesworthia* association is believed to have formed prior to *Glossina* radiation. In support, extensive concordant evolution has been observed between each tsetse species and its specific *Wigglesworthia* isolate (40). This ancient symbiont is localized intracellular in a bacteriome organ at the anterior midgut while an additional extracellular population is localized within the milk glands of females (29, 41). The functional roles of *Wigglesworthia* include nutrient provisioning, (42-44), contributing B-vitamins typically lacking in blood, and influencing host immunological (45, 46) and microbial community robustness (9, 10, 47). The importance of the symbiosis is demonstrated by the loss of female fecundity upon the removal of *Wigglesworthia* (22, 48, 49), which can partially be restored through the provisioning of B-vitamins (42) or yeast extract (22) in blood meals, providing strong evidence towards its nutrient contributory role. Additionally, the absence of *Wigglesworthia* during larval

stages results in immune-compromised adults with impaired development of the peritrophic matrix, separating the contents of the blood meal with the midgut epithelium, which ultimately influences the timing of immune response (50). These combined effects cause *Wigglesworthia*-free tsetse to be significantly more susceptible to microbial infections (45, 50, 51).

Despite the different co-evolutionary histories of the symbionts with the tsetse fly, adaptations resulting from co-residence, specifically the interdependency of metabolites, have commenced. While the provisioning of B-vitamins by *Wigglesworthia* is vital to host fitness (42), the production of thiamine (Vitamin B1) by this symbiont also influences *Sodalis* proliferation (9). Notably, *Sodalis* lacks the capability to produce thiamine, yet retains a functional ABC transporter (9) for the acquisition of exogenous sources of thiamine. These findings support a recent theory on reductive genome evolution known as the Black Queen Hypothesis, where the production of a 'leaky' product (in this case thiamine) may lead to the loss of that functional capability by a beneficiary, driving interspecies dependency (52). The idea of *Wigglesworthia* supplying public goods also brings into question their utilization by trypanosomes during development within tsetse, as they compensate for the lack of many biosynthetic pathways by encoding transporters to sequester metabolites (53).

The annotated genomes of *Wigglesworthia*, isolated from *Glossina morsitans* (Wgm) (44) and *G. brevipalpis* (Wgb) (43), share commonalities with other ancient symbionts (reviewed in (14)). These characteristics include extreme Adenine-Thymine (AT) bias and a highly reduced size (~ 700 kb), with the retention of only genes believed to be necessary to maintain the mutualism. These *Wigglesworthia* genomes demonstrate extensive chromosomal synteny, despite an ancient separation and subsequent host species co-diversification (44). Comparative genomic analyses between the two *Wigglesworthia* species enables the development of

hypotheses regarding functionality differences between these symbionts. Interestingly, one of the few features unique to Wgm, relative to the reference Wgb genome, is that it encodes the complete chorismate biosynthesis pathway (Fig. S1), which converts phosphoenolpyruvate (PEP) and erythrose 4-phosphate into chorismate, a precursor for the production of the aromatic amino acids and vitamins (54, 55). Wgm is then able to incorporate chorismate into the *p*-aminobenzoate (pABA) biosynthesis branch for downstream folate (Vitamin B9) production. Folate is required for all life, as it is involved in DNA synthesis, repair and methylation (56-58). These metabolome distinctions between the *Wigglesworthia* species are particularly striking, given the lack of chorismate and folate biosynthetic capabilities by *T. brucei* subspp. (folate transporters are retained within the genome) (53, 59) coupled with the higher vector competency of the Wgm host, *G. morsitans*, relative to its sister species, *G. brevipalpis* (32-36). The unique capabilities of *Wigglesworthia* species may aid in ecologically significant host phenotypic variation, as seen with other symbiotic systems (2, 60-63), which in contrast to the *Wigglesworthia* symbiosis, are much more recent in association.

Here, we aim to characterize the distinct retention of chorismate and folate biosynthetic potential by Wgm, to determine if natural selection has likely preserved these capabilities during host co-evolution. To accomplish this, we first performed expression analyses of various Wgm loci involved in chorismate and folate production, as well as the corresponding host transporter within bacteriomes, relative to host sex and development to determine whether these are actively and differentially transcribed. We next conducted detection assays within tsetse bacteriomes to confirm folate production and to determine if differences in symbiont biosynthetic loci expression directly corresponded with relative nutrient abundance. Furthermore, functional studies examined the role of Wgm chorismate and folate production towards various aspects of

G. morsitans biology; specifically life longevity, digestion, reproduction and trypanosome infection. Lastly, we aim to elucidate whether these unique pockets of loci between ancient mutualists contribute towards host phenotypic differences arising co-evolution.

MATERIALS AND METHODS

Tsetse flies. *G. morsitans morsitans* flies were maintained in the Department of Biology insectary at West Virginia University at 24 ± 1°C with 55% relative humidity on a 12 hr light/12 hr dark schedule. Tsetse were fed defibrinated bovine blood (Hemostat Laboratories Dixon, CA) every 48 hr using an artificial membrane feeding system (64). Trypanosome infection assays were performed at Yale University School of Public Health.

Reverse transcriptional analyses of Wgm chorismate and folate biosynthetic loci. To examine the expression of the Wgm chorismate and folate biosynthetic pathway loci (Fig. S1), the transcription of *aroA* (3-phosphoshikimate 1-carboxyvinyltransferase), *pabB* (aminodeoxychorismate synthase subunit I), and *folP* (7,8-dihydropteroate synthase) were assessed. Tsetse flies were sacrificed and bacteriomes were dissected and placed in RNA*later* (Ambion, Austin, TX), if not immediately processed. Total RNA was isolated following the TRIzol protocol (Invitrogen, Carlsbad, CA), treated with DNaseI (Ambion), and verified free of DNA contamination through PCR using RNA only as template. First-strand cDNA synthesis was performed with ~140 ng RNA, a 2 μM primer cocktail of gene specific 3' end primers (Table 1; aroAqPCRRev, pabBqPCRRev, folPqPCRRev, and rpsCqPCRRev) and Superscript II Reverse Transcriptase (Invitrogen). Quantitative PCR (qPCR) was performed with SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA), 0.4 mM gene specific primers (Table 1), and 2 μL cDNA template in a Bio-Rad CFX96 Real-Time PCR Detection System with 30 amplification cycles. The quantification of amplicons, relative to standards, was analyzed with Bio-Rad CFX Manager

software version 2.0 with transcript abundance normalized to Wgm *rpsC* (30S ribosomal protein S3) expression. The species-specificity of the primers used for reverse transcription was verified by using *Sodalis* culture DNA as a negative control. Each sample was analyzed in triplicate and averaged. Negative controls were included in all amplification reactions.

Folate detection within bacteriomes. A standard *Lactobacillus rhamnosus* microbiological assay (65) was used for folate quantification in wild-type (WT) and aposymbiotic (generated by obtaining progeny from females maintained on 25 μg/mL tetracycline + 1% (w/v) yeast extract supplemented blood (22, 47)) flies. Bacteriomes were dissected, individually placed into 0.1 M K_2HPO_4/KH_2PO_4 buffer + 1% ascorbic acid (pH 6.3) and immediately homogenized to release bacteriocyte contents. To release folate from bound proteins and denature folate degrading and interconversion enzymes, samples were placed in a H_2O bath at 100°C for 5 min. Subsequently, bacteriome samples were incubated with charcoal pre-treated rat plasma (Wistar Rat Plasma, Innovative Research, Novi, MI) serving as a conjugase of folate polyglutamates to monoglutamates at 37°C for 30 min. Samples were then sterilized using a 0.22 μm pore filter, and 1:100 dilutions were mixed with folate deficient Difco folic acid casei medium (Becton, Dickinson and Company, Sparks, MD) supplemented with 20 μg/mL chloramphenicol (Chl; Sigma, St. Louis, MO) and inoculated with log phase *L. rhamnosus*^{Chl} (ATCC # 27773) at an OD_{600} = 0.01. The cultures were incubated for 18 hr at 37°C. Standard concentrations of folic acid (Sigma) (10 – 125 fmol/well) were also mixed with *L. rhamnosus*^{Chl} culture to create a standard curve. A positive control of 300 fmol folic acid and negative control of initial buffer only were subjected to the complete procedure to ensure the retention of initial folate and lack of additional folate, respectively. Moreover, an additional negative control included only *L. rhamnosus*^{Chl}, lacking folate supplementation. The growth of *L. rhamnosus* was measured by a

Biomate3 spectrophotometer (Thermo Fisher Scientific, Madison WI) using absorbance readings at 600 nm. Folate content of bacteriome samples were determined by comparing to the standard curve (of which R^2 values of 0.95 – 0.99 were obtained) and expressed as pmol folate per bacteriome. At least 4 bacteriomes were sampled for each group.

Determination of bacteriome *Wigglesworthia* density. To differentiate whether changes in folate abundance in bacteriomes were due to either higher transcriptional levels or to an increase in symbiont density, tsetse were sacrificed and DNA was isolated from dissected bacteriomes using the Holmes-Bonner protocol (66). The density of *Wigglesworthia* was determined with qPCR as previously described (10).

Investigation of tsetse folate transporter expression within the bacteriome. Expression of the *G. morsitans* folate transporter (Aksoy, personal communication; Vectorbase gene ID: GMOY005445), known as a reduced folate carrier (RFC), was examined within bacteriome samples using semi-quantitative reverse transcriptional (RT-PCR) analysis. Bacteriome RNA was isolated and reverse transcription, using a 3' gene primer cocktail of GmBtubrev and GmRFCrev (Table 1), was performed as described above. Second strand synthesis was performed with complementary 5' end gene primers (Table 1) and 2 uL cDNA template for 35 amplification cycles. The products were analyzed using agarose gel electrophoresis and visualized with Kodak one image analysis software. Three samples were included for each time point examined. Negative controls were included in all PCR reactions. The constitutively expressed *G. morsitans* beta-tubulin gene (GenBank Accession number DQ377071) was used as a loading control and to verify RNA integrity.

Impact of trypanosome challenge on Wgm chorismate and folate biosynthetic loci expression. To examine the expression of genes involved in chorismate and folate biosynthesis

during trypanosome challenge, teneral *G. morsitans* received an initial blood meal that contained bloodstream *T. b. brucei* RUMP503 parasites (2 x 10^6/ml). Subsequently, flies were maintained on defibrinated bovine blood meals. At 2 and 5 wks of age, flies were dissected and midguts and salivary glands (5 wks only) were microscopically scored for trypanosome infections. Bacteriome RNA was isolated from infected, refractory and sex and age-matched non-challenged controls and reverse transcriptional analyses of Wgm *aroA, pabB* and *folP* were performed as described above.

Impact of chorismate pathway inhibition on *G. morsitans* biology. Different concentrations of glyphosate (*N*-(phosphonomethyl)glycine) (Sigma), a specific competitive enzymatic inhibitor of AroA (67-69), were administered by incorporation into *G. morsitans* blood meals. The impact of symbiont chorismate pathway inhibition towards tsetse life longevity, digestion, and reproduction were examined.

Longevity. Survival was monitored from both WT and aposymbiotic adult flies (maintained on tetracycline treatments throughout adulthood) to differentiate between the effects of glyphosate treatment towards tsetse or symbiont biology. Flies were maintained on diet combinations consisting of blood only or supplemented with 10 or 20 mM glyphosate. Dead flies were recorded daily for a duration of sixty days.

Digestion. Flies were maintained on 0 or 10 mM glyphosate-supplemented blood meals. At 2 weeks of age, flies were offered their respective blood meals for 20 min and only those that had fed (confirmed through visual inspection of bloody abdomens) remained in the study. Blood meal digestion was determined by measuring undigested hemoglobin levels using the cyanmethemoglobin method (Sigma), as previously described (48). Briefly, tsetse midguts (n=3 per time point and treatment) were dissected, homogenized in 2 mL Drabkin's reagent (Sigma),

and incubated at ambient temperature in the dark for 15 min. Absorbance measurements were taken of each sample at 540 nm. Hemoglobin concentration was determined by comparing the mean absorbance reading of an unknown to a standard curve (R^2=0.98) created using bovine hemoglobin (Sigma) and expressed as mg/gut.

Tsetse reproductive output and progeny development. To examine whether symbiont chorismate pathway inhibition would impact larval development, virgin females were maintained on defibrinated bovine blood only or supplemented with 10 mM glyphosate and/or 500 nM folic acid and mated with WT males after 2 feedings. At 20 d old, the reproductive tracts (including the spermatheca, oocytes, and uterus) were removed using a Leica S6D dissection microscope (Leica Microsystems, Heerbrugg, Switzerland) and image captured with a Leica DFC420 digital color camera and Leica Application Suite Version 2.8.0 software. Measurements of the uterus, containing a developing larva if pregnant, were then obtained using ImageJ (70) and compared to age-matched WT virgin and mated females. To examine the effect of chorismate inhibition towards pupal deposition and eclosion, virgin females were maintained on the supplemented blood meals described above and mated with WT males after 2 feedings. Pupae deposition and maternal fly mortality were monitored for 45 days (~ 2 gonotrophic cycles). Pupal weights, eclosion rates, and teneral wing area (obtained using ImageJ) used as a measure of body size (71, 72), were also recorded.

Statistical analyses. Data were analyzed using Microsoft Excel for Student's *t*-test or Mann-Whitney U-tests, with comparison of variances determined using F-tests. JMP 7.0 (SAS Institute, Cary, NC, USA) was used to perform ANOVAs, Chi-Square tests, and survival analyses. Survival curves were created using the Kaplan-Meier method and compared using the log rank test (73).

RESULTS

Wgm chorismate and folate biosynthetic loci exhibit differential expression between tsetse sex, during development and pregnancy. To investigate Wgm chorismate and folate biosynthetic gene expression within bacteriomes relative to host sex, development and during reproduction, the transcript abundance of *aroA*, *pabB,* and *folP* were determined by qPCR (Fig. 1). Comparison of 2 wk old tsetse flies revealed higher *aroA* and *pabB* transcript levels within the bacteriomes of virgin females, although statistical significance was only observed with the former (Student *t*-test; $p = 0.02$), with similar *folP* expression between sexes (Fig. 1A). To further investigate the role of Wgm chorismate and folate provisioning during female adulthood and pregnancy, the transcriptional profile of these loci from the bacteriomes of 5 wk old virgin and similarly aged adult flies during late gestation (i.e. carrying a 3rd instar larva) were also examined. The expression of all three loci was significantly higher in the bacteriomes of pregnant flies (Fig. 1B, Mann-Whitney U-test; *aroA* $p = 0.01$ and *pabB* $p = 0.02$; Student's *t*-test; *folP* $p = 0.03$) upon comparison with similarly aged virgin females. Additionally, Wgm *folP* transcript abundance was also significantly higher in pregnant females, compared to 2 wk virgins (Fig 1B, Student's *t*-test; $p = 0.05$). Further investigation into chorismate biosynthesis loci expression during gonotrophic cycles revealed that there is variation between pregnancies, with higher pathway expression during the second gonotrophic cycle relative to the first (Fig. S2). These data not only indicate that chorismate and folate biosynthesis loci are actively transcribed by Wgm, but also supports a more significant functional role for these loci towards female host biology, specifically during early adulthood and pregnancy.

Bacteriome folate content is highest during early female adulthood and pregnancy. To move beyond gene expression analyses, a standard *L. rhamnosus* microbiological assay was

employed to enable both the detection and quantification of folate abundance from single bacteriome isolates. Comparison of the bacteriomes within 2 wk male and virgin female adults demonstrate that there are sex-specific differences in folate content, with virgin female bacteriomes containing significantly more folate (Fig. 2A, Student's *t*-test; $p = 0.016$). Interestingly, the transcriptional analyses of 2 wk female bacteriomes demonstrated increased expression of Wgm *aroA* and *pabB* (Fig. 1A) compared to 2 wk males. To determine whether higher folate content in 2 wk virgin females resulted from increased transcription or differences in symbiont density, comparisons of the Wgm bacteriome populations were performed with qPCR. No significant difference in Wgm density between 2 wk males and virgin females (Student's *t*-test; $p = 0.196$, data not shown) was observed. Previous research has reported that at 2 wks, virgin female tsetse harbor a significantly greater population of Wgm than males (74), yet this study examined whole fly symbiont densities, which includes the female milk gland Wgm population that is believed to proliferate at this time in development (Rio, personal communication). Therefore, folate content in 2 wk virgin females bacteriomes is likely due to an increase in transcriptional activity, rather than heightened symbiont densities. Notably, folate was not detected in the bacteriome samples from 2 wk old aposymbiotic male and virgin female flies (Fig. 2A), supporting the production of this nutrient exclusively by Wgm within bacteriomes. To further explore folate production during reproduction, age-matched pregnant and virgin female bacteriome folate contents were also compared, revealing a significantly higher abundance of folate harbored within the bacteriomes of pregnant individuals (Fig. 2A, Student's *t*-test; $p = 0.0001$). In support of previous research, which has shown that mating does not influence symbiont population density within females (74), Wgm densities were found to be comparable within 3 wk virgin and mated bacteriomes (Student's *t*-test; $p = 0.311$, data not

shown). Therefore, the independent observations of greater Wgm chorismate and folate biosynthesis loci transcript abundance (Fig. 1A and 1B) as well as higher folate quantities (Fig. 2A) from the bacteriomes of young adult and pregnant females, indicates that Wgm folate production is more significant during these times.

***G. morsitans* folate transporter is expressed within bacteriomes.** To examine the activity of the *G. morsitans* RFC within the bacteriome, semi-quantitative reverse transcriptional analysis was performed and revealed that expression of this transporter is dynamic within adult tsetse (Fig. 2B). Transcript abundance of *G. morsitans* RFC directly corresponded with the amount of folate detected within the bacteriomes of the respective samples, which was highest in pregnant females (Fig. 2A).

Trypanosome challenge influences Wgm chorismate and folate biosynthesis loci transcription. The impact of *T. b. brucei* RUMP503 challenge on Wgm chorismate and folate biosynthetic loci expression was examined at two stages of infection and compared with age- and sex-matched non-challenged controls. Early stage infection (2 wks post infectious meal) was microscopically determined by parasite presence in the midgut while late-stage infection (5 wks post infectious meal) was identified by observation of both midgut and salivary gland colonization by trypanosomes. Although bacteriomes isolated from refractory females at 2 wks following trypanosome challenge had higher transcriptional activity of *aroA*, *pabB,* and *folP*, no statistically significant differences were observed between females (Fig. 3A). Contrastingly in males at 2 wks, bacteriomes of infected individuals had significantly higher Wgm *aroA* (Mann-Whitney U-test; $p = 0.05$) and *pabB* (Mann-Whitney U-test; $p = 0.044$) expression compared to non-challenged individuals (Fig. 3B). Infected males also had higher *pabB* (Student's *t*-test; $p = 0.033$) and *folP* (Student's *t*-test; $p = 0.05$) transcript levels when compared to refractory flies

(Fig. 3B). At 5 wks post infectious meal, all challenged females irrespective of infection status exhibited higher transcript abundance relative to age-matched non-challenged controls (Fig. 3C). Infected individuals had significantly higher bacteriome Wgm *aroA* (Student's *t*-test; $p = 0.028$), *pabB* (Student's *t*-test; $p = 0.002$), and *folP* (Student's *t*-test; $p = 0.015$) expression levels compared to non-challenged females. The Wgm loci examined also exhibited higher expression in refractory, compared to non-challenged controls; yet statistical significance was only obtained for *pabB* (Mann-Whitney U-test; $p = 0.03$) (Fig. 3C). A disparate transcriptional profile was observed in male tsetse at 5 wks, upon comparison with females, with a significant increase in transcript abundance of Wgm *pabB* and *folP* in bacteriomes from refractory flies, compared to those infected (Fig. 3D, Student's *t*-test $p = 0.006$ and Mann-Whitney U-test $p = 0.019$, respectively). Moreover, Wgm *pabB* expression was also significantly higher in refractory in comparison to within age-matched non-challenged male bacteriomes (Student's *t*-test; $p = 0.046$). Distinctions in Wgm chorismate and folate biosynthetic loci expression during trypanosome challenge may be a contributing factor to the variation observed in vector competency among sexes of *G. morsitans* (30).

Chorismate pathway inhibition impacts tsetse biology. As Wgm appears to be actively synthesizing folate within bacteriomes, with higher production during early female adulthood and pregnancy, we further examined the role of this nutrient provisioning towards various tsetse fitness parameters, specifically, tsetse life longevity, blood meal digestion and reproduction, by enzymatically inhibiting chorismate biosynthesis through blood meal glyphosate supplementation.

Host life longevity. There have been conflicting reports as to the toxic effects of glyphosate on animals (75-79). Glyphosate inhibits chorismate production by competitively binding to AroA

(3-phosphoshikimate 1-carboxyvinyltransferase), preventing the incorporation of PEP (69) and 3-phosphoshikimate to create 5-O-(1-carboxyvinyl)-3-phosphoshikimate, which is subsequently converted into chorismate. While Wgm encodes the chorismate and folate biosynthesis pathways, searches of respective enzymatic components within the currently in progress *G. morsitans* genome (80) did not produce any homologous hits, supporting the belief that animals generally lack this capability (54, 81). Therefore, glyphosate should directly impact symbiont rather than tsetse fly biology, with any negative consequences resulting from treatment likely alleviated by decreasing symbiont populations. To differentiate between a direct toxic effect of glyphosate on the tsetse host versus symbiont biology, we examined the survival of tsetse maintained on glyphosate (10 and 20 mM) and/or tetracycline (25 μg/mL) supplemented blood. Previously, tetracycline treatments have been shown to nearly clear symbionts within directly impacted individuals (45, 48). Interestingly, tetracycline only treatments significantly increased tsetse longevity (males, log rank test; $p < 0.0001$ and females, log rank test; $p < 0.0001$) (Fig. 4), possibly due to the reduction in cost required for typically maintaining the symbiotic association. The supplementation of 10 mM glyphosate did not have adverse effects on life longevity, as the survival curves were similar to the controls, fed blood only (males, log rank test; $p = 0.7547$ and females, log rank test; $p = 0.9463$) (Fig. 4). Female tsetse administered tetracycline and 10 mM glyphosate had significantly higher survival rates compared to those fed blood alone (log rank test; $p = 0.0001$) or 10 mM glyphosate supplementation (log rank test; $p < 0.0001$). Although not statistically significant, a similar trend was observed in male flies maintained on both tetracycline and 10 mM glyphosate supplemented blood compared to those fed 10 mM glyphosate (log rank test; $p = 0.4612$) and blood alone (log rank test; $p = 0.2057$) (Fig. 4). The increased survival of tsetse maintained on both tetracycline and glyphosate, compared to only

glyphosate, suggests that this compound mostly impacts symbiont biology. In contrast to 10 mM glyphosate, 20 mM glyphosate supplementation did result in a life longevity cost in WT flies (Fig. S3), suggesting that there may be adverse, toxic effects towards the host at this high of a concentration. Tsetse maintained on 20 mM glyphosate supplemented blood had significantly shorter survival than controls (males, log rank test; $p < 0.0001$ and females, log rank test; $p < 0.0001$) (Fig. S3). As 10 mM glyphosate does not appear to be toxic towards the host with respect to life longevity, this concentration was used to investigate the inhibition of chorismate biosynthesis on other aspects of tsetse biology.

Digestion. During the survival study, visual inspection of tsetse abdomens indicated that individuals administered glyphosate had prolonged blood meal digestion (Fig. S4). Therefore, we further examined the impact of glyphosate (10 mM) on digestion by measuring midgut hemoglobin levels of treated 2 wk old flies. Interestingly, flies maintained on blood alone took larger blood meals, as represented by higher hemoglobin levels with abdomens 1 hr post feeding, although statistical significance was only observed in males (Fig. 5, Student's *t*-test; $p = 0.001$) (female data not shown). Despite taking smaller blood meals, males maintained on 10 mM glyphosate contained significantly more hemoglobin 48 hr post feeding within their midguts (Student's *t*-test; $p = 0.01$), indicating compromised digestive capabilities, with a similar trend observed in females although lacking significance.

Host reproduction. To examine whether chorismate pathway inhibition, and consequential suppression of downstream folate production, influences intrauterine larval development, mated females were maintained on specific blood meal regimens (Fig. 6A) and the reproductive tracts were dissected at 20 days post emergence; a time point associated with late intrauterine larval development during the 1st gonotrophic cycle within WT females (82). The mean intrauterine

area of the group maintained on 10 mM glyphosate supplemented blood was significantly smaller than mated controls (Fig. 6A, Mann-Whitney U-test; $p = 0.05$) and similar in size to virgins, suggesting that chorismate biosynthesis is critical for larval development. Notably, the bacteriome folate content, determined using the *L. rhamnosus* microbiological assay, from age-matched, mated females maintained on 10 mM glyphosate supplemented blood was significantly lower than those fed blood alone (Student's *t*-test; $p = 0.02$) (Fig. 6B). These data demonstrate that glyphosate supplementation in the blood impacts Wgm chorismate production within bacteriocytes, resulting in decreased downstream folate production that significantly hampers *in utero* larval development. When 500 nM folic acid was used to rescue the reproductive impact of glyphosate administration, there was an increase in intrauterine area, which was significantly larger than virgins (Mann-Whitney U-test; $p = 0.05$) and similar to mated controls, further supporting the downstream utilization of chorismate for folate biosynthesis by Wgm and the importance of folate for larval development. Folic acid supplementation alone resulted in larger uterine areas than age-matched WT flies, indicating a role for this nutrient during pregnancy. These data demonstrate that females maintained on glyphosate do remain fecund, but that inhibition of the chorismate pathway negatively impacts larval development.

To further investigate the necessity of Wgm produced folate in tsetse reproduction, flies were maintained on the supplemented blood meals described above and larval deposition was observed for 45 days. There were no significant differences in time to first larval deposition (ANOVA; $p = 0.28$), larval deposition per female (ANOVA; $p = 0.9$) or pupal eclosion (Chi-Square test; $p = 0.7$) between the treatment groups over the observed period (data not shown). Pupal weight (Fig. 6C) and wing areas of teneral progeny (Fig. 6D) from females maintained on blood only or 500 nM folic acid were significantly greater than those deposited by mothers fed

10 mM glyphosate +/- 500 nM folic acid. Interestingly, progeny from the 10 mM glyphosate treatment had a significantly longer pupal life stage than the blood only or 10 mM glyphosate + 500 nM folic acid groups for both sexes (Fig. 6E), possibly accommodating for the deposition of larvae that were underweight (Fig. 6C) and may not have been fully developed. These results coupled with the intrauterine area measurements suggest that Wgm folate production is important for the developmental progression of tsetse progeny. In addition to a significantly longer pupal period, the progeny from females administered glyphosate were physically smaller, in regards to both pupal weight and teneral wing area, showing that lengthening the developmental time does not recover offspring size.

DISCUSSION

Ancient bacterial symbionts from closely related host species have highly similar genomic content, yet small pockets of unique loci have been noted (44, 83). The preservation of these symbiont metabolome distinctions through evolutionary time, particularly given the strong reductive genome processes encountered by these symbionts (14), suggests that they may have been selected for and play a vital part in the biology and evolution of their host species. Here, we use the tsetse fly, specifically *G. morsitans*, and its obligate mutualist, Wgm, to investigate the significance of one small set of symbiont-species specific genes towards host biology. Of the CDSs encoded within the Wgm genome, absent in the Wgb reference genome, approximately 40% are involved in chorismate biosynthesis and its downstream incorporation into folate production. Although these represent a relatively small number of genes, given the significance of chorismate and folate as precursors in thymidylate (dTMP), purine and amino acid synthesis, and their pivotal role in biological processes, such as DNA production and cell growth (56, 57, 81), the retention of these loci result in distinctive metabolic capabilities by this symbiont species

(44). Here, we demonstrate that the Wgm chorismate and subsequent folate biosynthesis pathways not only have remained functional, but also maintain significance towards tsetse host fitness.

These studies provide the first piece of evidence supporting the actual production of a nutrient (i.e. folate) within the bacteriome by this ancient obligate mutualist and a potential means of dissemination through a synchronous host transporter. The transcriptional profile of Wgm chorismate and folate biosynthetic loci between sexes and during female development and pregnancy corresponds with the relative amounts of folate detected within these bacteriomes, appearing highest during pregnancy. Within aposymbiotic flies the amount of folate was below the level of detection, providing additional evidence for the production of this nutrient, within the bacteriome, solely by the Wgm symbiont. These findings further highlight the importance of this mutualism for tsetse host reproduction, as Wgm contributes essential nutrients aiding in the growth and development of intrauterine larva (22, 42-44). Additionally, the annotation of the *G. morsitans* genome resulted in identification of a single folate transporter, known as a RFC (Aksoy, personal communication). Studies performed with other animal models have shown that the RFC maintains bidirectional properties (84), that is, it can be used for both import and export of specific nutrients. The expression of the *G. morsitans* RFC observed within bacteriome samples suggests that this transporter is present in bacteriocytes and may be used to disseminate Wgm produced folate, although further investigation is required.

To explore the role of symbiont chorismate and downstream folate production on host biology, flies were administered glyphosate, a chorismate biosynthesis pathway inhibitor, within their blood meals. Historically, there have been conflicting reports on whether glyphosate directly impacts insects or animals. According to the glyphosate technical report composed by

the National Pesticide Information Center (85), no toxic effects on honey bees were found and the greatest threat of glyphosate to arthropods was alterations to habitat and food availability (86). Research with *Drosophila* reported that glyphosate caused genotoxic effects, using the wing spot test (75). Those results contradict an earlier study, which tested the genotoxic potential of glyphosate using multiple microbial and animal *in vivo* and *in vitro* tests, where no genotoxic activity was observed (76). This current study provides evidence that a high concentration of glyphosate (20 mM) may be toxic to tsetse flies, as life longevity was significantly reduced and survival was not influenced by symbiont presence, despite the target of this compound being absent within the host genome. In contrast, administration of a lower concentration of glyphosate (10 mM) does not impact longevity, compared to those fed blood only (Fig. 4). At this concentration, there also appears to be symbiont-specific effects, as symbiont absence increases host survival (Fig. 4) and glyphosate treatment resulted in decreased folate levels within bacteriomes (Fig. 6B). Therefore, we demonstrate that while 20 mM glyphosate supplemented to tsetse via blood meals may be toxic, impacting the host directly, decreasing the concentration alleviates these effects, while still inhibiting symbiont chorismate production.

In support of the hypothesis that these unique Wgm loci retain importance for tsetse biology, inhibiting symbiont chorismate production negatively impacted the host. First, chorismate biosynthesis pathway inhibition resulted in decreased digestion, which is similar to a previous study examining the biology of Wgm-free flies (48), further elucidating the role for this symbiont in host blood meal digestion. In *Drosophila*, digestion has been associated with midgut cell turnover (87). The inhibited digestive capabilities seen in tsetse maintained on 10 mM glyphosate may be a consequence of cell turnover impairment, as folate is required for cell generation and is not being supplied by the symbiont, yet additional studies should be performed

to further explore this hypothesis. Negative effects of glyphosate treatment may be due to decreased nutrient availability, but could also be a result of the accumulation of intermediates (specifically shikimate-3-phosphate), as this compound inhibits the action of AroA, likely contributing to the bloodier abdomens through time of treated tsetse individuals. Glyphosate treatment also resulted in decreased Wgm folate production, negatively impacting intrauterine larval development. Although the larvae from glyphosate-treated females were deposited at the same rate as the WT group, the pupae were smaller and had a significantly longer pupal stage. The reduced progeny size and longer developmental time may have negative impacts toward their biology, elongating the already slow reproductive cycle of tsetse and possibly leading to additional fitness costs. These findings correspond with past research demonstrating the importance of folate in the diet of insects for successful reproduction and growth (88, 89). Interestingly, antifolates administered within the diet of the buffalo fly (*Haematobia irritans exigua*), which feeds on blood, resulted in adult female sterility and longevity costs (90). Folate analogues (methotrexate and aminopterin) have also been shown to inhibit larval development and induce sterility of female insects (90-92), which can be regained by supplementation of exogenous folate (92). Similar to the results of this study, increasing the concentration of folic acid within the diet of the housefly, *Musca domestica L.*, also caused an increase in larval growth (92). This study contributes to the body of work demonstrating the importance of folate for all animals during reproduction and further elucidates a role for bacterial symbionts in its production.

While the *Sodalis* genome also retains the capability to produce chorismate and folate, Wgm appears to be primarily responsible for this nutrient provisioning. Evidence includes past research, which showed that flies lacking Wgm but still retaining *Sodalis* are reproductively

sterile (48), demonstrating that *Sodalis* cannot compensate for the decrease in provisioned nutrients within the fly. The absence of Wgm through tsetse host generations (capable of reproducing when maintained on yeast supplemented blood meals) also resulted in the eventual loss of *Sodalis* (47), supporting a metabolic dependence by *Sodalis* on Wgm, which has previously been shown with thiamine (9). Additionally, within the field, *Sodalis* is not present in all tsetse populations or individuals (25, 93), indicating that this secondary symbiont has established more recently and does not play as pivotal of a role in the holobiont success. To ensure that this study examined the contributions of Wgm only, folate production was measured within the bacteriome organ, which is densely packed with Wgm cells, and also utilized primers that specifically amplified Wgm loci. Future studies should investigate the *Sodalis* chorismate and folate biosynthetic pathways to determine whether they retain functionality or are in the process of being removed from the genome. In support of the later hypothesis, the *Sodalis* genome is undergoing drastic reduction, encoding a high number of pseudogenes (94, 95), during the course of adapting to a symbiotic lifestyle (9, 96, 97). Additional research may also examine whether these loci retain more significant roles in tsetse species whose *Wigglesworthia* symbionts have lost these functional capabilities, leading to further genomic complementation of biosynthetic capabilities as reported with ancient co-resident symbionts (5, 7, 8, 98).

In addition to metabolic provisioning by Wgm to the tsetse host and *Sodalis*, there is a growing body of evidence demonstrating a role for Wgm in host immune development and trypanosome refractoriness (45, 46, 48, 50, 51, 99). The transcriptional activity of the Wgm chorismate and folate biosynthetic loci differs during trypanosome challenge between sex and stages of infection. At 2 wks after an infectious blood meal, there is sufficient time for trypanosomes to have successfully established or been cured from the tsetse midgut, and if

infected at 5 wks of age, flies can also harbor mature salivary gland infections and are now capable of transmission to naïve hosts. The increase in Wgm chorismate and folate biosynthesis loci expression in 2 wk refractory females, as well as in 5 wk challenged females and refractory male flies, may indicate a role for Wgm folate provisioning in repairing damaged host cells resulting from an increased humoral immune response, required by the fly to clear the infection. Bloodstream form trypanosomes are susceptible to reactive oxygen species (ROS) (100, 101) and have been shown to cause a significant increase in H_2O_2 levels within tsetse hosts (102). Additionally, blocking ROS in midguts by feeding L-glutathione significantly increases midgut infection rates (30), further supporting the role of ROS in clearing early infections. ROS produced to fight infection can also cause self-inflicted cell damage, which may be balanced with midgut epithelial cell repair (103), aided by nutrient provisioning from Wgm. In support, *G. morsitans* with self-cleared *T. b. brucei* infections exhibited a greater amount of up-regulated peroxidase homologs within the midgut, when compared to those that remained infected (38).

The variation in expression profiles of Wgm *aroA*, *pabB* and *folP* may be a contributing factor to the sex-specific vector competency differences observed (30, 31). Immune stimulation due to trypanosome infection has been shown to result in a fitness cost (104). Increased transcription of Wgm chorismate and folate biosynthetic loci in challenged females may be a response by the symbiont to heightened immune stimulation implemented to clear infections, as previously seen upon bacterial challenge (105), to enhance longevity and provide the greatest chance for reproductive success. Additionally, the higher transcription of these Wgm loci at early infection in males may reflect an initial increase in Wgm nutrient production, required by the parasitized host. Male hosts have a significantly higher transmission index than females (30), possibly due to the faster migration from the midgut to the salivary glands via the foregut (30).

Importantly, trypanosomes are unable to produce folate and encode transporters to salvage this nutrient from its surroundings (53), which is essential for *T. brucei* thymidylate synthesis and growth (106). The supplementation of nutrients upon early infection, produced by Wgm, may inadvertently be a beneficial resource for trypanosomes that have successfully established within the midgut, contributing to faster developmental progression within these flies. The data from this study supports the hypothesis that Wgm influences the vector competency of their host and also may contribute to the host's ability to recover from the damages incurred by humoral immune activation, yet warrants further investigation.

As folate is essential for all life (56-58), the absence of these pathways within the Wgb genome (43) brings into question how *G. brevipalpis* obtains this required nutrient. The loss of the chorismate and folate biosynthesis capabilities may have initially resulted from stochastic mutations in pathway genes due to high levels of genetic drift coupled with extreme population bottlenecks (14). Once the biosynthetic pathways were no longer intact, additional loci may have then purged from the genome. Due to the concordant evolution of *Wigglesworthia* and its tsetse host species (40), changes in ecological factors between tsetse species may have resulted from the loss or retention of specific symbiont capabilities. One adaptation may have been vertebrate host blood meal preference, which differs between *G. morsitans* and *G. brevipalpis* (107-109), possibly accommodating for the lack of folate production by Wgb. Folate content in blood has been shown to vary between animals (110, 111) and is largely dependent on diet (111). Additionally, while *G. brevipalpis* is known to have lower vector competency than *G. morsitans* (32-36), the contributing factors resulting in this phenotype remain unknown but may include these symbiont metabolome distinctions. The retention of capabilities within specific

Wigglesworthia species, such as folate production, may have influenced evolutionary and phenotypic variation within *Glossina*.

This study describes the metabolic interdependence of Wgm chorismate and folate production towards *G. morsitans* biology. Here, we show that the retention of these nutrient biosynthesis capabilities by Wgm retain functionality and appear to be important for host biology, specifically during pregnancy and trypanosome challenge. The loss of chorismate and folate biosynthetic capabilities within Wgb may have been due to random genetic drift, possibly influencing the ecology of the tsetse host. The evolution of symbiont genomes, consequently impacting their metabolic capabilities, may therefore contribute to the evolution of phenotypic differences among associated host species.

REFERENCES

1. Sullivan JT, Patrick HN, Lowther WL, Scott DB, & Ronson CW (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci USA 92(19):8985-8989.
2. Kikuchi Y, et al. (2012) Symbiont-mediated insecticide resistance. Proc Natl Acad Sci USA 109(22):8618-8622.
3. Kikuchi Y & Graf J (2007) Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. Appl Environ Microbiol 73(6):1984-1991.
4. Bomar L, Maltz M, Colston S, & Graf J (2011) Directed culturing of microorganisms using metatranscriptomics. mBio 2(2):e00012-00011.
5. McCutcheon JP, McDonald BR, & Moran NA (2009) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci USA 106(36):15394-15399.
6. McCutcheon JP & Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol 2:708-718.
7. McCutcheon JP & von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr Biol 21(16):1366-1372.
8. Wu D, et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol 4(6):e188.
9. Snyder AK, Deberry JW, Runyen-Janecky L, & Rio RVM (2010) Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc Biol Sci 277(1692):2389-2397.

10. Snyder AK, McLain C, & Rio RVM (2012) The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density via exogenous nutrient provisioning. Appl Environ Microbiol 78(21):7792-7797.

11. Rio RVM, Lefevre C, Heddi A, & Aksoy S (2003) Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. Appl Environ Microbiol 69(11):6825-6832.

12. Nyholm SV, Stewart JJ, Ruby EG, & McFall-Ngai MJ (2009) Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. Environ Microbiol 11(2):483-493.

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

14. McCutcheon JP & Moran NA (2011) Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10(1):13-26.

15. Lopez-Madrigal S, Latorre A, Porcar M, Moya A, & Gil R (2011) Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. J Bacteriol 193(19):5587-5588.

16. Scholle MD & Gerdes SY (2008) Whole-genome detection of conditionally essential and dispensable genes in *Escherichia coli* via genetic footprinting. Methods Mol Biol 416:83-102.

17. Lukjancenko O, Wassenaar TM, & Ussery DW (2010) Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol 60(4):708-720.

18. Oh S S, Buddenborg S, Yoder-Himes DR, Tiedje JM, & Konstantinidis KT (2012) Genomic diversity of *Escherichia* isolates from diverse habitats. PLoS One 7(10):e47005.

19. Aksoy S (1995) *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. Int J Syst Bacteriol 45(4):848-851.

20. Dale C & Maudlin I (1999) *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int J Syst Bacteriol 49:267-275.

21. Zhou W, Rousset F, & O'Neill S (1998) Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc Biol Sci 265(1395):509-515.

22. Alam U, et al. (2011) *Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans*. PLoS Pathog 7(12):e1002415.

23. Geiger A, et al. (2009) First isolation of *Enterobacter*, *Enterococcus*, and *Acinetobacter* spp. as inhabitants of the tsetse fly (*Glossina palpalis palpalis*) midgut. Infect Genet Evol 9(6):1364-1370.

24. Geiger A, Fardeau ML, Falsen E, Ollivier B, & Cuny G (2010) *Serratia glossinae* sp. nov., isolated from the midgut of the tsetse fly *Glossina palpalis gambiensis*. Int J Syst Evol Microbiol 60:1261-1265.

25. Lindh JM & Lehane MJ (2011) The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie Van Leeuwenhoek 99(3):711-720.

26. Leak S (1999) Tsetse biology and ecology. Nairobi, Kenya: CABI Publishing.

27. Tobe SS & Langley PA (1978) Reproductive physiology of *Glossina*. Ann Rev Entomol 23:283-307.

28. Ma WC & Denlinger DL (1974) Secretory discharge and microflora of milk gland in tsetse flies. Nature 247:301-303.

29. Balmand S, Lohs C, Aksoy S, & Heddi A (2013) Tissue distribution and transmission routes for the tsetse fly endosymbionts. J Invertebr Pathol 112 Suppl:S116-122.

30. Peacock L, Ferris V, Bailey M, & Gibson W (2012) The influence of sex and fly species on the development of trypanosomes in tsetse flies. PLoS Negl Trop Dis 6(2):e1515.

31. Dale C, Welburn SC, Maudlin I, & Milligan PJ (1995) The kinetics of maturation of trypanosome infections in tsetse. Parasitology 111:187-191.

32. Harley JM (1971) Comparison of the susceptibility of infection with *Trypanosoma rhodesiense* of *Glossina pallidipes*, *G. morsitans*, *G. fuscipes* and *G. brevipalpis*. Ann Trop Med Parasitol 65(2):185-189.

33. Moloo SK & Kutuza SB (1988) Comparative study on the susceptibility of different *Glossina* species to *Trypanosoma brucei brucei* infection. Trop Med Parasitol 39(3):211-213.

34. Moloo SK, Kabata JM, & Sabwa CL (1994) A study on the maturation of procyclic *Trypanosoma brucei brucei* in *Glossina morsitans centralis* and *G. brevipalpis*. Med Vet Entomol 8(4):369-374.

35. Moloo SK, Okumu IO, & Kuria NM (1998) Comparative susceptibility of *Glossina longipennis* and *G. brevipalpis* to pathogenic species of *Trypanosoma*. Med Vet Entomol 12(2):211-214.

36. Moloo SK, Orinda GO, Sabwa CL, Minja SH, & Masake RA (1999) Study on the sequential tsetse-transmitted *Trypanosoma congolense*, *T. brucei brucei* and *T. vivax*

infections to African buffalo, eland, waterbuck, N'Dama and Boran cattle. Vet Parasitol 80(3):197-213.

37. Hao Z, et al. (2001) Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc Natl Acad Sci USA 98(22):12648-12653.

38. Lehane MJ, et al. (2003) Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. Genome Biol 4(10):R63.

39. Nayduch D & Aksoy S (2007) Refractoriness in tsetse flies (Diptera: Glossinidae) may be a matter of timing. J Med Entomol 44(4):660-665.

40. Chen X, Li S, & Aksoy S (1999) 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(1):49-58.

41. Cheng Q & Aksoy S (1999) Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. Insect Mol Biol 8(1):125-132.

42. Nogge G (1981) Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology 82:101-104.

43. Akman L, et al. (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet 32(3):402-407.

44. Rio RVM, et al. (2012) Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. mBio 3(1):e00240-11.

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

46. Weiss B & Aksoy S (2011) Microbiome influences on insect host vector competence. Trends Parasitol 27(11):514-522.

47. Wang J, Brelsfoard C, Wu Y, & Aksoy S (2013) Intercommunity effects on microbiome and GpSGHV density regulation in tsetse flies. J Invertebr Pathol 112 Suppl:S32-39.

48. Pais R, Lohs C, Wu Y, Wang J, & Aksoy S (2008) The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Appl Environ Microbiol 74(19):5965-5974.

49. Nogge G (1976) Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. Experientia 32(8):995-996.

50. Weiss BL, Wang J, Maltz MA, Wu Y, & Aksoy S (2013) Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. PLoS Pathog 9(4):e1003318.

51. Weiss BL, Maltz M, & Aksoy S (2012) Obligate symbionts activate immune system development in the tsetse fly. J Immunol 188(7):3395-3403.

52. Morris JJ, Lenski RE, & Zinser ER (2012) The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio 3(2):e00036-00012.

53. Berriman M, et al. (2005) The genome of the African trypanosome *Trypanosoma brucei*. Science 309(5733):416-422.

54. Herrmann KM & Weaver LM (1999) The Shikimate Pathway. Annu Rev Plant Physiol Plant Mol Biol 50:473-503.

55. Dosselaere F & Vanderleyden J (2001) A metabolic node in action: chorismate-utilizing enzymes in microorganisms. Crit Rev Microbiol 27(2):75-131.

56. Koury MJ & Ponka P (2004) New insights into erythropoiesis: the roles of folate, vitamin B12, and iron. Annu Rev Nutr 24:105-131.

57. Lucock M (2004) Is folic acid the ultimate functional food component for disease prevention? BMJ 328(7433):211-214.

58. Suzuki H, Toyota M, Sato H, Sonoda T, Sakauchi F, Mori M. (2006) Roles and causes of abnormal DNA methylation in gastrointestinal cancers. Asian Pac J Cancer Prev 7(2):177-185.

59. Jackson AP, et al. (2010) The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human african trypanosomiasis. PLoS Negl Trop Dis 4(4):e658.

60. Moran NA (2007) Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci USA 104:8627-8633.

61. Oliver KM, Degnan PH, Hunter MS, & Moran NA (2009) Bacteriophages encode factors required for protection in a symbiotic mutualism. Science 325(5943):992-994.

62. Janson EM, Peeden ER, Stireman JO, & Abbot P (2010) Symbiont-mediated phenotypic variation without co-evolution in an insect-fungus association. J Evol Biol 23(10):2212-2228.

63. Tsuchida T, et al. (2010) Symbiotic bacterium modifies aphid body color. Science 330(6007):1102-1104.

64. Moloo SK (1971) An artificial feeding technique for Glossina. Parasitology 63(3):507-512.

65. Arcot J & Shrestha A (2005) Folate: methods of analysis. Trends Food Sci Tech 16(6-7):253-266.

66. Holmes DS & Bonner J (1973) Preparation, molecular-weight, base composition, and secondary structure of giant nuclear ribonucleic-acid. Biochemistry 12(12):2330-2338.

67. Steinrucken HC & Amrhein N (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic-acid-3-phosphate synthase. Biochem Biophys Res Commun 94(4):1207-1212.

68. Boocock MR & Coggins JR (1983) Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate. FEBS Lett 154(1):127-133.

69. Steinrucken HC & Amrhein N (1984) 5-enolpyruvylshikimate-3-phosphate synthase of *Klebsiella pneumoniae* 2. Inhibition by glyphosate [N-(phosphonomethyl)glycine]. Eur J Biochem 143(2):351-357.

70. Rasband WS (2012) ImageJ (U.S. National Institutes of Health, Bethesda, MD) http://imagej.nih.gov/ij/.

71. Robertson FW (1959) Studies in quantitative inheritance. Xii. Cell size and number in relation to genetic and environmental variation of body size in *Drosophila*. Genetics 44(5):869-896.

72. Carron A (2007) Correlation between wing measurements and dry body weight in male and female *Ochlerotatus* (*Ochlerotatus*) *caspius* (Pallas, 1771) (Diptera: Culicidae). European Mosquito Bulletin 24:4-8.

73. Bewick V, Cheek L, & Ball J (2004) Statistics review 12: survival analysis. Crit Care 8(5):389-394.

74. Rio RVM, Wu YN, Filardo G, & Aksoy S (2006) Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. Proc Biol Sci 273(1588):805-814.

75. Kaya B, Creus A, Yanikoglu A, Cabre O, & Marcos R (2000) Use of the *Drosophila* wing spot test in the genotoxicity testing of different herbicides. Environ Mol Mutagen 36(1):40-46.

76. Li AP & Long TJ (1988) An evaluation of the genotoxic potential of glyphosate. Fundam Appl Toxicol 10(3):537-546.

77. Lipok J (2009) Dual action of phosphonate herbicides in plants affected by herbivore--model study on black bean aphid *Aphis fabae* rearing on broad bean *Vicia faba* plants. Ecotoxicol Environ Saf 72(6):1701-1706.

78. Schneider MI, Sanchez N, Pineda S, Chi H, & Ronco A (2009) Impact of glyphosate on the development, fertility and demography of *Chrysoperla externa* (Neuroptera: Chrysopidae): ecological approach. Chemosphere 76(10):1451-1455.

79. Evans SC, Shaw EM, & Rypstra AL (2010) Exposure to a glyphosate-based herbicide affects agrobiont predatory arthropod behaviour and long-term survival. Ecotoxicology 19(7):1249-1257.

80. Megy K, et al. (2012) VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res 40:D729-734.

81. de Crecy-Lagard V, El Yacoubi B, de la Garza RD, Noiriel A, & Hanson AD (2007) Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genomics 8:245.

82. Attardo GM, Guz N, Strickler-Dinglasan P, & Aksoy S (2006) Molecular aspects of viviparous reproductive biology of the tsetse fly (*Glossina morsitans morsitans*): regulation of yolk and milk gland protein synthesis. J Insect Physiol 52(11-12):1128-1136.

83. van Ham RC, et al. (2003) Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci USA 100(2):581-586.

84. Matherly LH & Goldman DI (2003) Membrane transport of folates. Vitam Horm 66:403-456.

85. National Pesticide Information Center (2010) Glyphosate Technical Fact Sheet. (Oregon State University, Corvallis, OR), pp 1-14.

86. Giesy JP, Dobson S, Solomon KR (2000) Ecotoxicological Risk Assessment for Roundup® Herbicide. Rev Environ Contam Toxicol 167:35-120.

87. Ohlstein B & Spradling A (2006) The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. Nature 439(7075):470-474.

88. Fraenkel G & Blewett M (1946) Folic acid in the nutrition of certain insects. Nature 157:697.

89. Fraenkel G & Blewett M (1947) The importance of folic acid and unidentified members of the vitamin B complex in the nutrition of certain insects. Biochem J 41(3):469-475.

90. Elvin CM, Liyou NE, Pearson R, Kemp DH, & Dixon NE (2003) Molecular cloning and expression of the dihydrofolate reductase (DHFR) gene from adult buffalo fly (*Haematobia irritans exigua*): effects of antifolates. Insect Mol Biol 12(2):173-183.

91. Affleck JG, Neumann K, Wong L, & Walker VK (2006) The effects of methotrexate on *Drosophila* development, female fecundity, and gene expression. Toxicol Sci 89(2):495-503.

92. Perry AS & Miller S (1965) The essential role of folic acid and the effect of antimetabolites on growth and metamorphosis of housefly larvae *Musca domestica* L. J Insect Physiol 11(9):1277-1287.

93. Farikou O, et al. (2010) 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(1):115-121.

94. Toh H, et al. (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res 16(2):149-156.

95. Belda E, Moya A, Bentley S, & Silva FJ (2010) Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics 11:449.

96. Weiss BL, Wu Y, Schwank JJ, Tolwinski NS, & Aksoy S (2008) An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. Proc Natl Acad Sci USA 105(39):15088-15093.

97. Maltz MA, Weiss BL, O'Neill M, Wu Y, & Aksoy S (2012) OmpA-mediated biofilm formation is essential for the commensal bacterium *Sodalis glossinidius* to colonize the tsetse fly gut. Appl Environ Microbiol 78(21):7760-7768.

98. McCutcheon JP & Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci USA 104(49):19392-19397.

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

100. Mehlotra RK (1996) Antioxidant defense mechanisms in parasitic protozoa. Crit Rev Microbiol 22(4):295-314.

101. Gobert AP, et al. (1998) Murine macrophages use oxygen- and nitric oxide-dependent mechanisms to synthesize S-nitroso-albumin and to kill extracellular trypanosomes. Infect Immun 66(9):4068-4072.

102. Hao Z, Kasumba I, & Aksoy S (2003) Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). Insect Biochem Mol Biol 33(11):1155-1164.

103. Welchman DP, Aksoy S, Jiggins F, & Lemaitre B (2009) Insect immunity: from pattern recognition to symbiont-mediated host defense. Cell Host Microbe 6(2):107-114.

104. Hu C, et al. (2008) Infections with immunogenic trypanosomes reduce tsetse reproductive fitness: potential impact of different parasite strains on vector population structure. PLoS Negl Trop Dis 2(3):e192.

105. Kaaya GP & Darji N (1988) The humoral defense system in tsetse: differences in response due to age, sex and antigen types. Dev Comp Immunol 12(2):255-268.

106. Sienkiewicz N, Jaroslawski S, Wyllie S, & Fairlamb AH (2008) Chemical and genetic validation of dihydrofolate reductase-thymidylate synthase as a drug target in African trypanosomes. Mol Microbiol 69(2):520-533.

107. FAO (1992) Training Manual for Tsetse Control Personnel, ed Pollock JN (Rome, Italy).

108. Clausen PH, et al. (1998) Host preferences of tsetse (Diptera: Glossinidae) based on bloodmeal identifications. Med Vet Entomol 12(2):169-180.

109. Farikou O, et al. (2010) Tsetse fly blood meal modification and trypanosome identification in two sleeping sickness foci in the forest of southern Cameroon. Acta Trop 116(1):81-88.

110. Schweigert BS & Pearson PB (1947) The folic acid content of blood from various species. Am J Physiol 148(2):319-322.

111. Leamon CP, et al. (2008) Impact of high and low folate diets on tissue folate receptor levels and antitumor responses toward folate-drug conjugates. J Pharmacol Exp Ther 327(3):918-925.

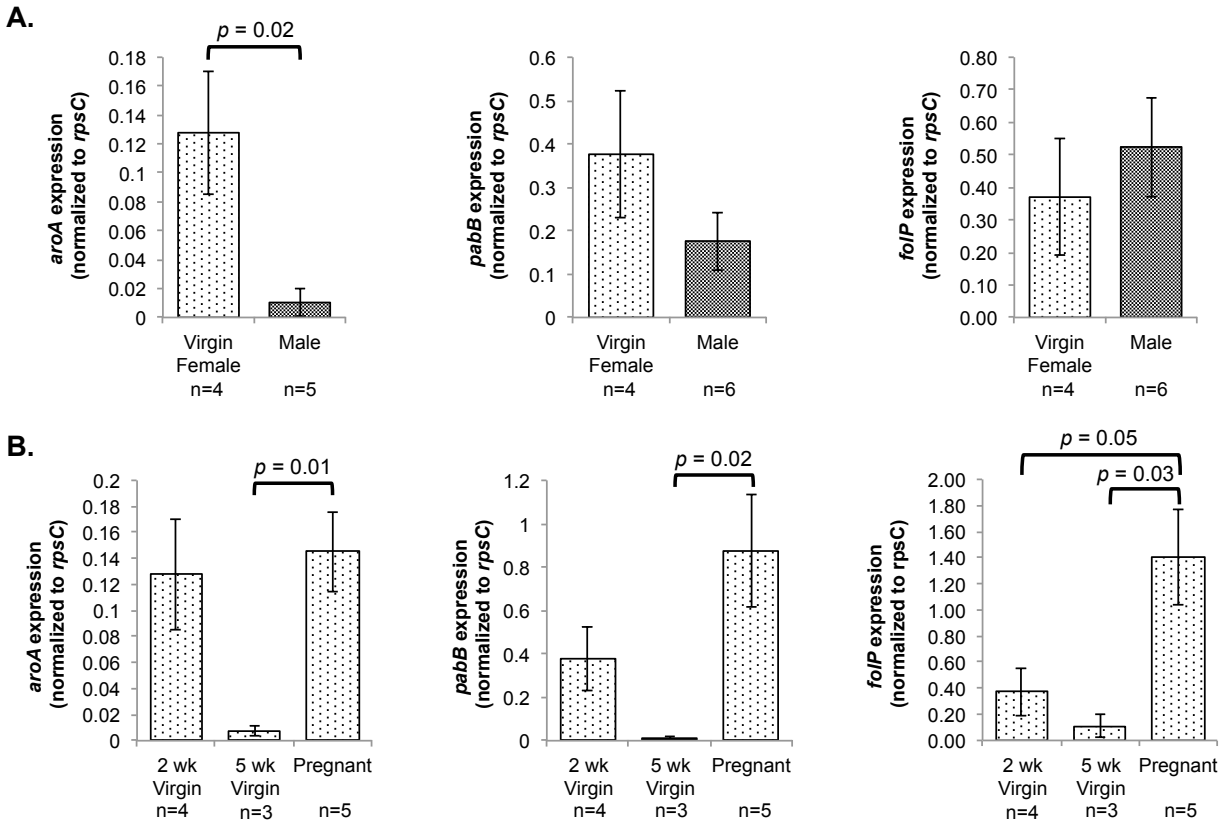


Figure 3-1. Wgm chorismate and folate biosynthetic loci exhibit differential expression within bacteriomes between (A.) tsetse sex at 2 wks old and (B.) through female adulthood and pregnancy. Graphs represent normalized transcript abundance with error bars signifying 1 standard error of the mean (S.E.M.). Student's *t*-tests and Mann-Whitney U-tests were performed when variances were equal and unequal, respectively, with statistically significant differences indicated above bars. Sample size (n) is indicated below each group.

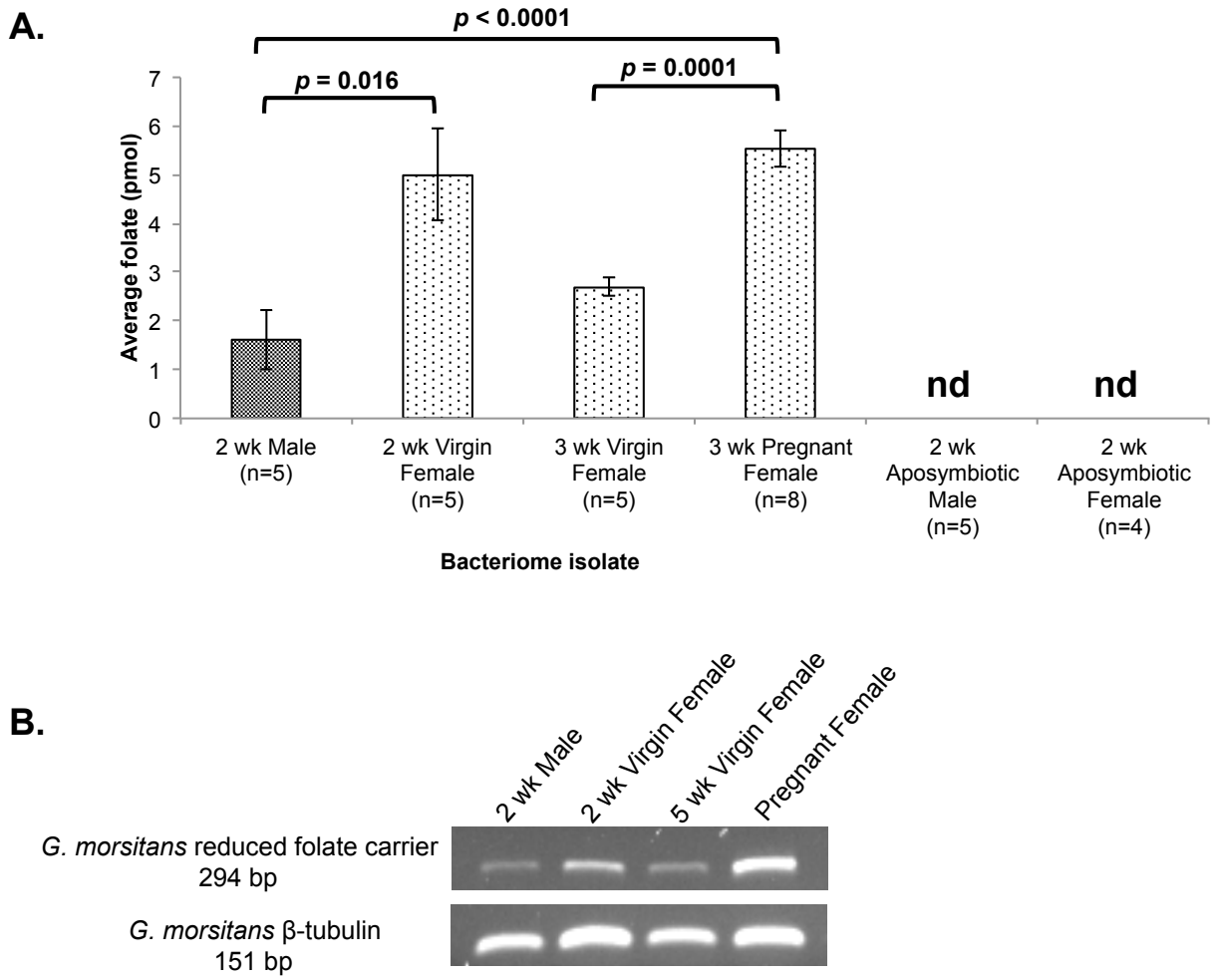


Figure 3-2. Folate content and host transporter expression within *G. morsitans* bacteriomes. (A.) Total folate quantity was determined using an *L. rhamnosus* microbiological assay. Significant differences between groups, determined using a Student's *t*-test, are indicated above bars. Error bars represent 1 S.E.M. Sample size (n) is indicated below each group. Samples with no folate detection are denoted as (nd). (B.) Semi-quantitative RT-PCR analysis of *G. morsitans* RFC (Reduced Folate Transporter) expression, with β-tubulin serving as a loading control. Representative samples are shown for each time point.

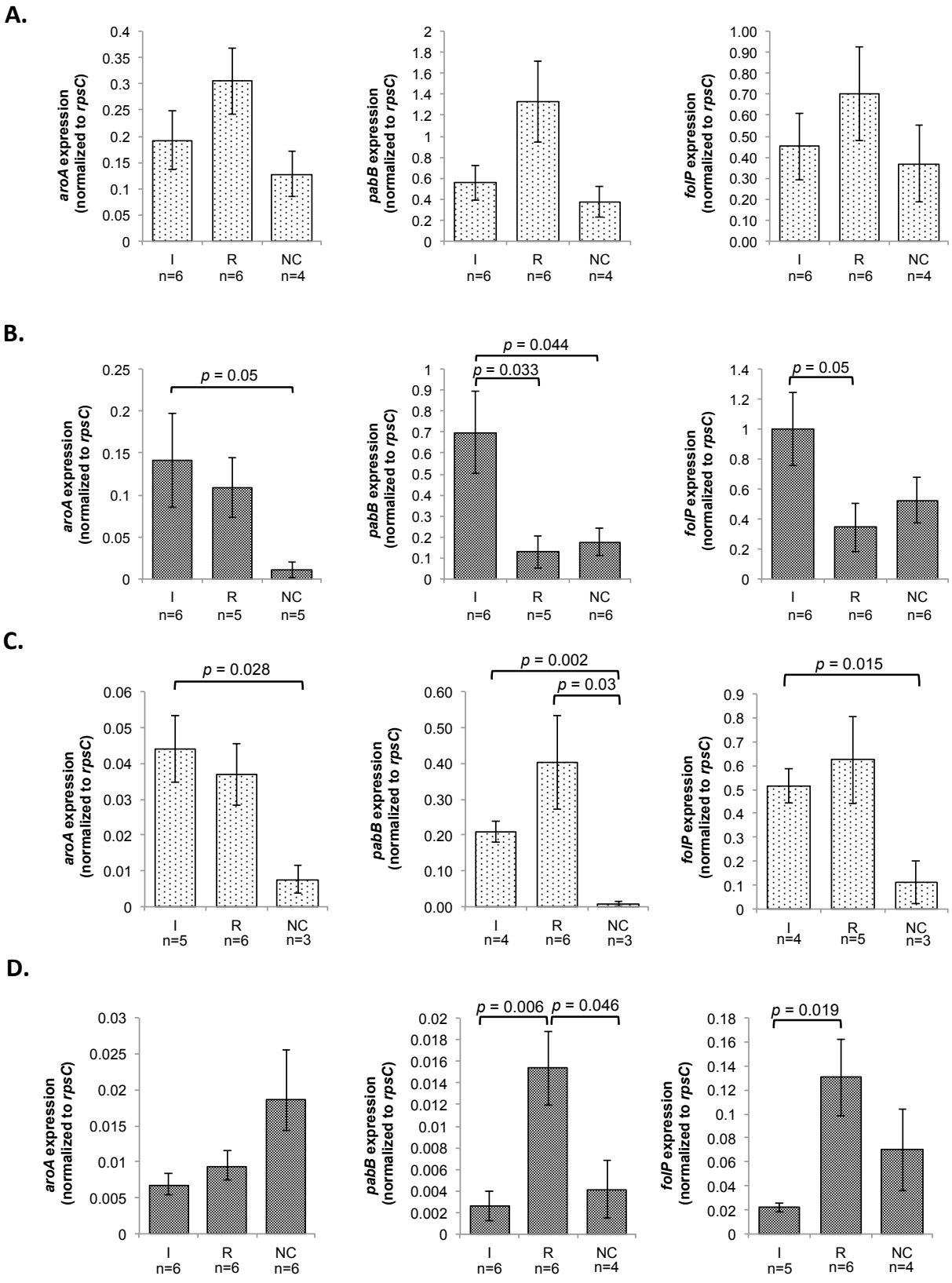

A.
aroA expression (normalized to rpsC)
pabB expression (normalized to rpsC)
folP expression (normalized to rpsC)
I n=6
R n=6
NC n=4
B.
p = 0.05
p = 0.044
p = 0.033
I n=6
R n=5
NC n=5
I n=6
R n=5
NC n=6
I n=6
R n=6
NC n=6
C.
p = 0.028
p = 0.002
p = 0.03
p = 0.015
I n=5
R n=6
NC n=3
I n=4
R n=6
NC n=3
I n=4
R n=5
NC n=3
D.
p = 0.006
p = 0.046
p = 0.019
I n=6
R n=6
NC n=6
I n=6
R n=6
NC n=4
I n=5
R n=6
NC n=4

Figure 3-3. Wgm chorismate and folate biosynthetic loci expression from the bacteriomes of 2 wk (A.) virgin female and (B.) male and 5 wk (C.) virgin female and (D.) male tsetse from *T. b. brucei* RUMP503 challenged and age-matched control flies. Infection status of midgut (2 wks), in addition to salivary glands (5 wks) is indicated as I; infected or R; refractory and compared to NC; non-challenged age-matched individuals. Loci expression was compared, using a Student's *t*-test or Mann-Whitney U-test, and significant differences are indicated above bars. Sample sizes (n) are indicated below respective bars. Standard error bars signify 1 S.E.M.

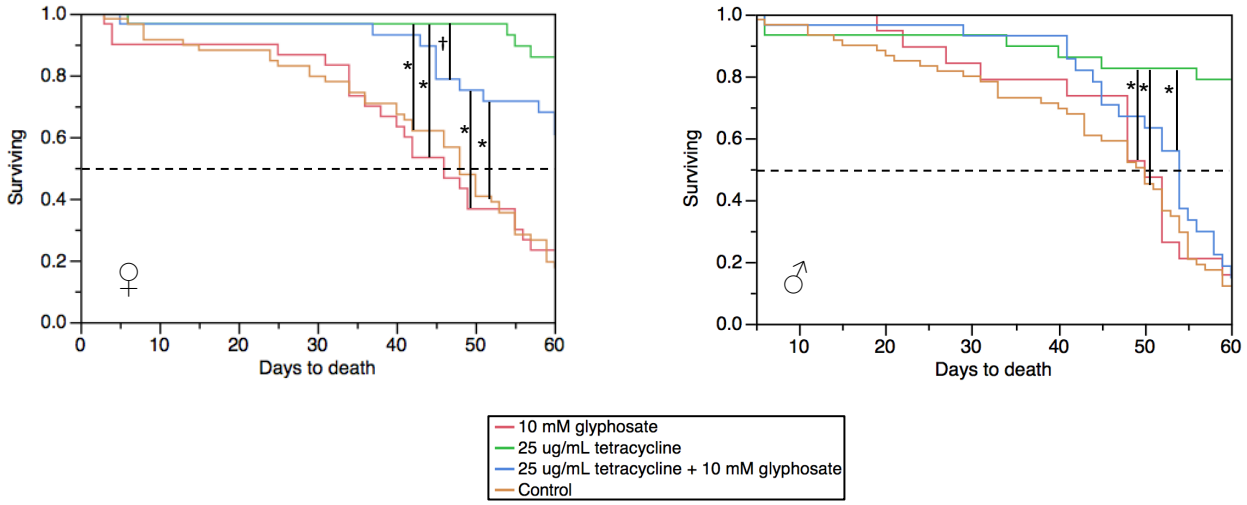


Figure 3-4. Survival curves, created in JMP 7.0 using the Kaplan-Meier method, of WT or aposymbiotic *G. morsitans* maintained on 10 mM glyphosate supplemented blood for 60 days. n ≥ 30 individuals per treatment. Significant differences between treatment groups, determined using the log-rank test, are indicated († $p = 0.035$, $^{*}p \leq 0.0001$).

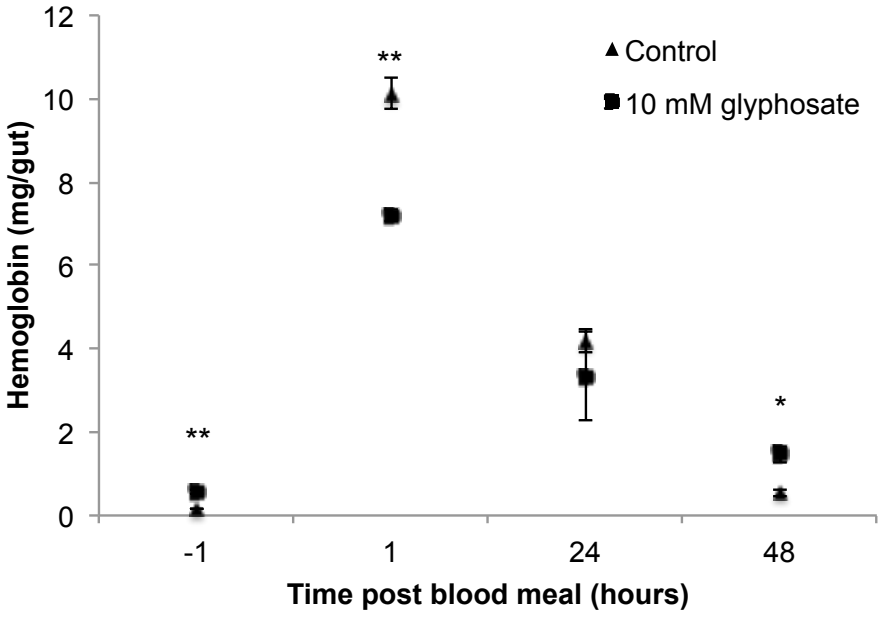
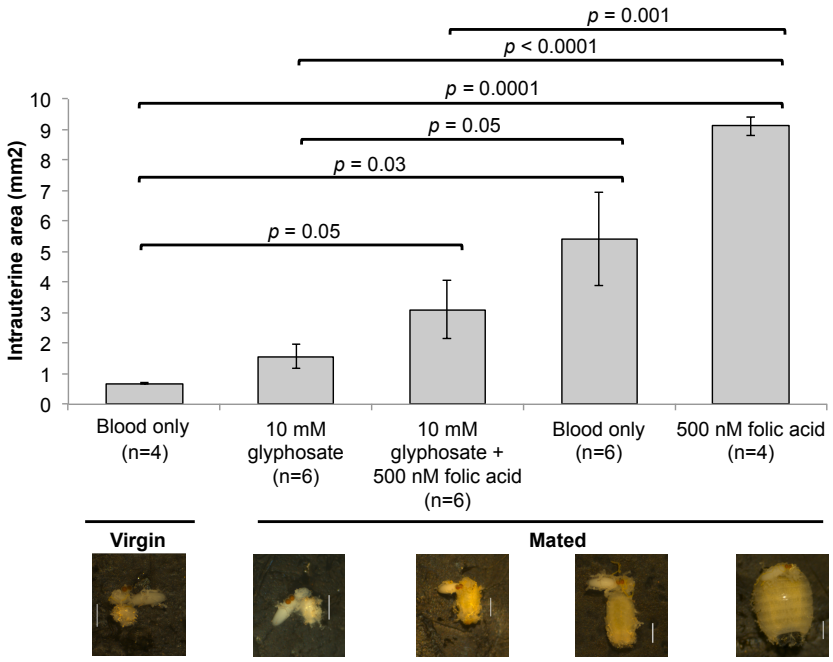
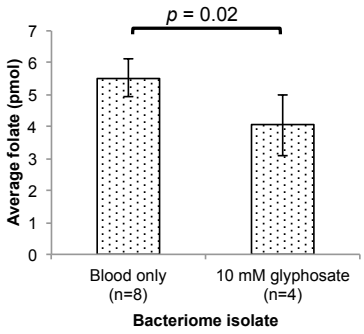


Figure 3-5. Hemoglobin concentration, representing undigested blood, from 2 wk old male *G. morsitans*. Significant differences between control and glyphosate treated flies at each time point are represented by * ($p < 0.05$) and ** ($p < 0.01$). Error bars indicate 1 S.E.M. n = 3 per time point and treatment group.

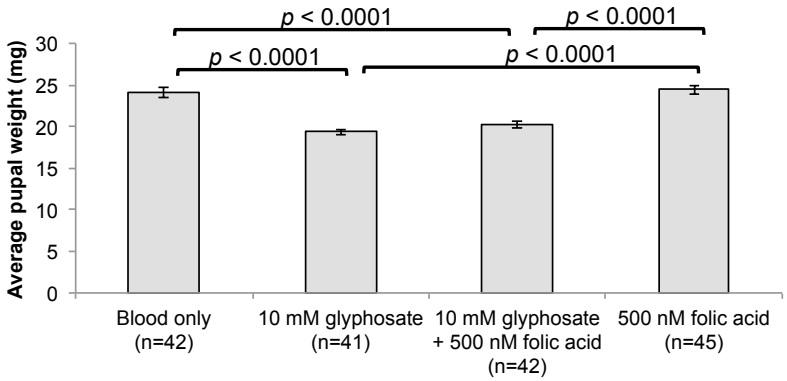
A.


p = 0.001
p < 0.0001
p = 0.0001
p = 0.05
p = 0.03
p = 0.05
Intrauterine area (mm2)
0
1
2
3
4
5
6
7
8
9
10
Blood only (n=4)
10 mM glyphosate (n=6)
10 mM glyphosate + 500 nM folic acid (n=6)
Blood only (n=6)
500 nM folic acid (n=4)
Virgin
Mated

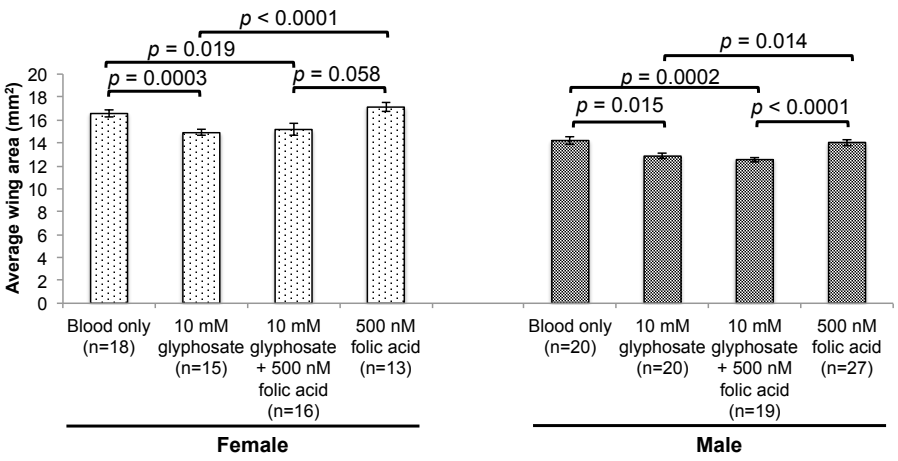
B.


p = 0.02
Average folate (pmol)
0
1
2
3
4
5
6
7
Blood only (n=8)
10 mM glyphosate (n=4)
Bacteriome isolate

C.


p < 0.0001
p < 0.0001
p < 0.0001
p < 0.0001
Average pupal weight (mg)
0
5
10
15
20
25
30
Blood only (n=42)
10 mM glyphosate (n=41)
10 mM glyphosate + 500 nM folic acid (n=42)
500 nM folic acid (n=45)

D.



E.

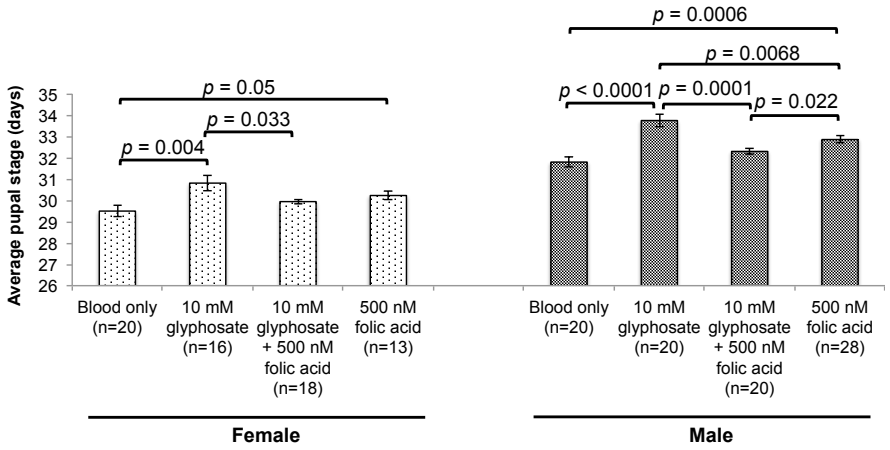


Figure 3-6. Inhibition of Wgm chorismate biosynthesis impacts progeny development. (A.) Comparison of age-matched (20 d) female uterine area (containing developing larva) maintained on indicated treatments. Representative images are included below each treatment group. (B.) Folate content within bacteriomes of 3 wk mated females maintained on blood ± 10 mM glyphosate supplementation. The (C.) pupal weight, (D.) teneral wing area, and (E.) length of the pupal life stage of progeny deposited from females maintained on specified blood meal supplementation. Statistically significant differences, determined using Student's *t*- or Mann-Whitney U-tests are indicated above bars. Error bars signify 1 S.E.M. n = sample size.

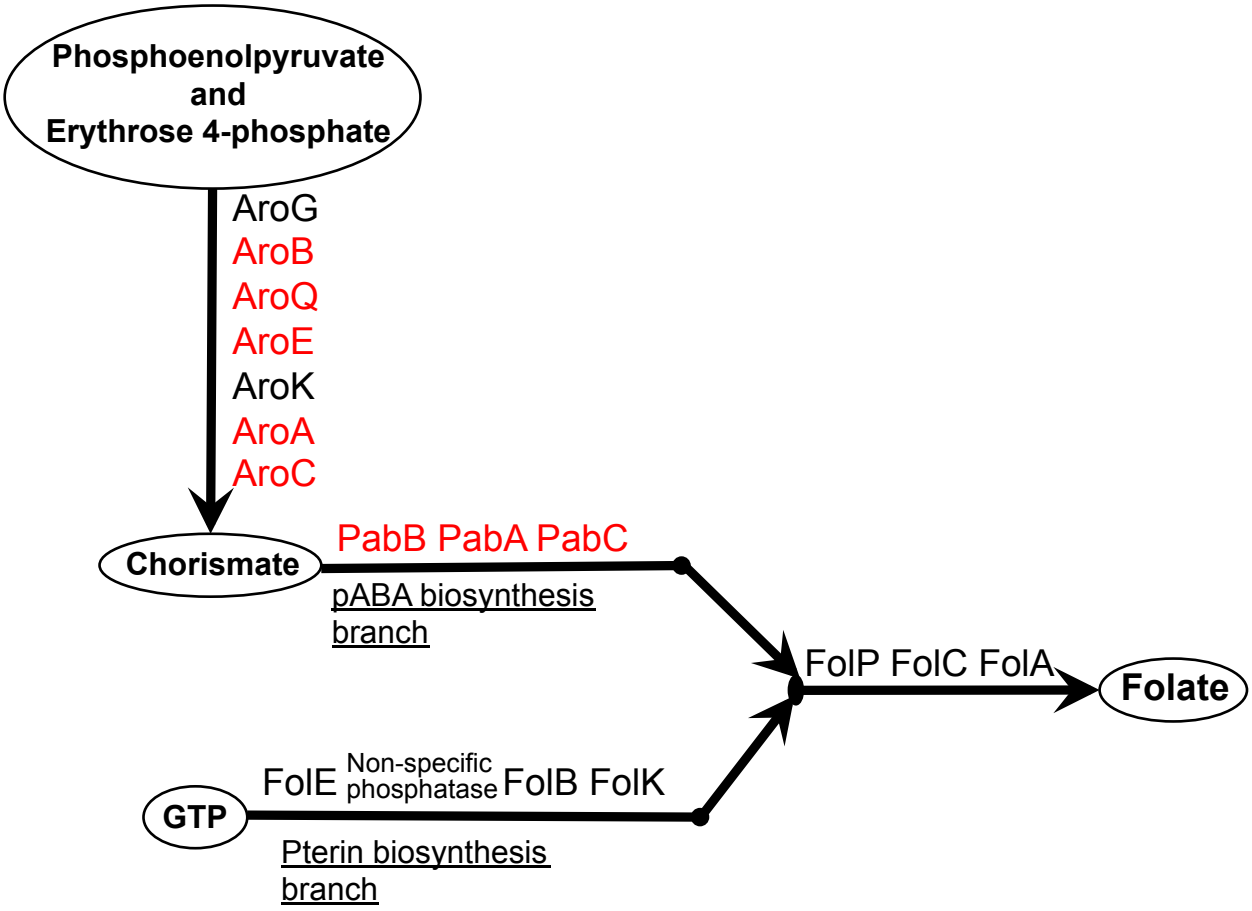


Figure 3-S1. Wgm retains the complete chorismate and folate biosynthesis pathways. The sequential chorismate and folate biosynthetic pathways are represented by arrows, indicating steps catalyzed by the enzymes named. Enzymes in red are only retained within Wgm, while those in black are also found in Wgb.

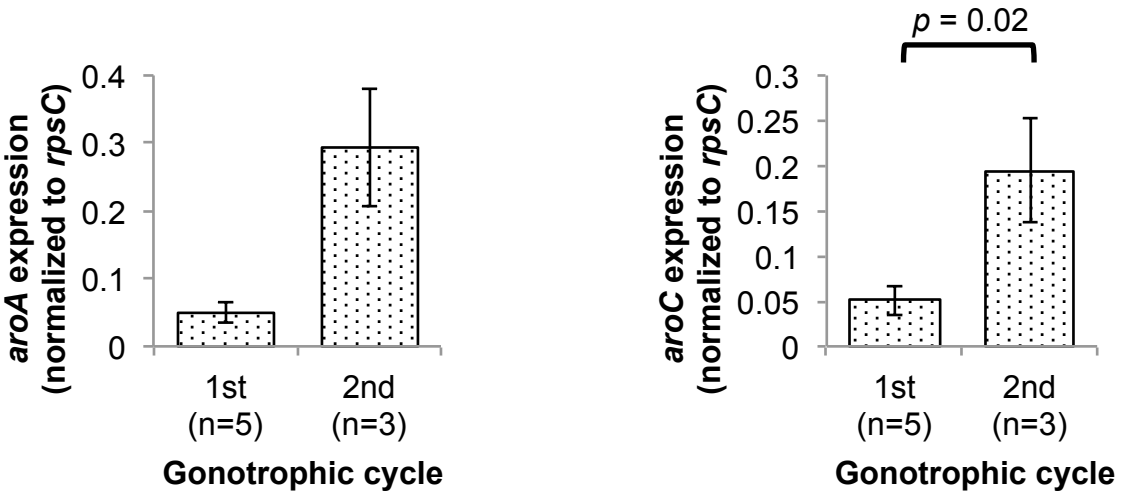


Figure 3-S2. Wgm chorismate biosynthesis loci expression differs between gonotrophic cycles. Graphs represent normalized transcript abundance with error bars signifying 1 S.E.M. In addition to *aroA* expression, the last step in chorismate biosynthesis, *aroC,* was also examined using Wgm-specific qPCR primers (aroCqPCRFor: 5' - GCA GTG AAA GGT ATT GAA ATT GG – 3' and aroCqPCRRev: 5' - AAG CAG GTT TAA TGG CAA GAG – 3') with a Ta = 50.5 °C. Significant differences are indicated above bars. n = sample size.

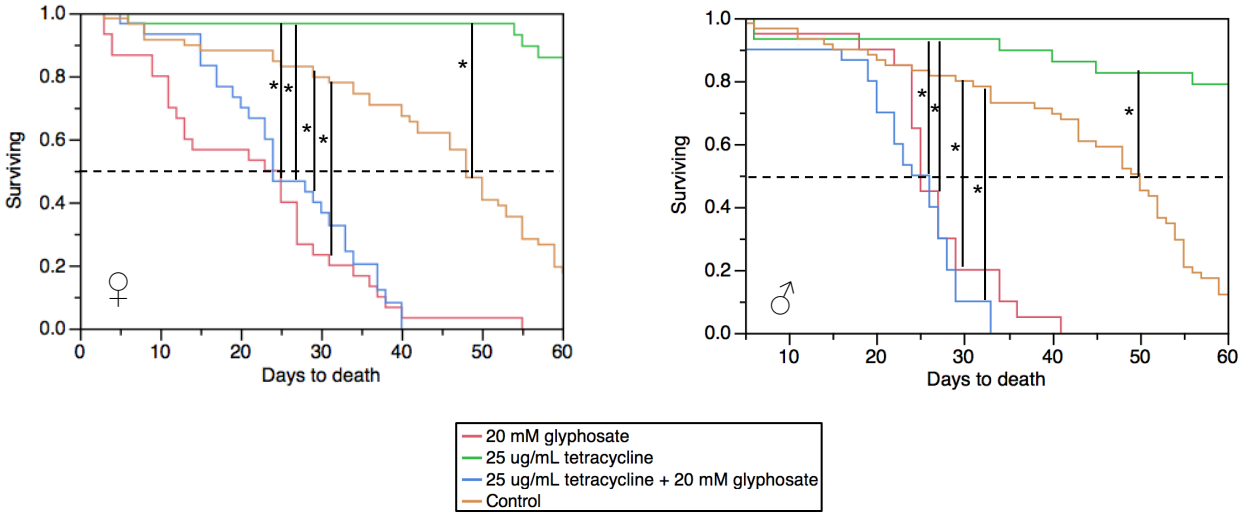


Figure 3-S3. Survival curves, created in JMP 7.0 using the Kaplan-Meier method, of WT or aposymbiotic *G. morsitans* maintained on 20 mM glyphosate supplemented blood for 60 days. n ≥ 30 individuals per treatment. Significant differences between treatment groups, determined using the log-rank test, are indicated ($^{*}p \leq 0.0001$).

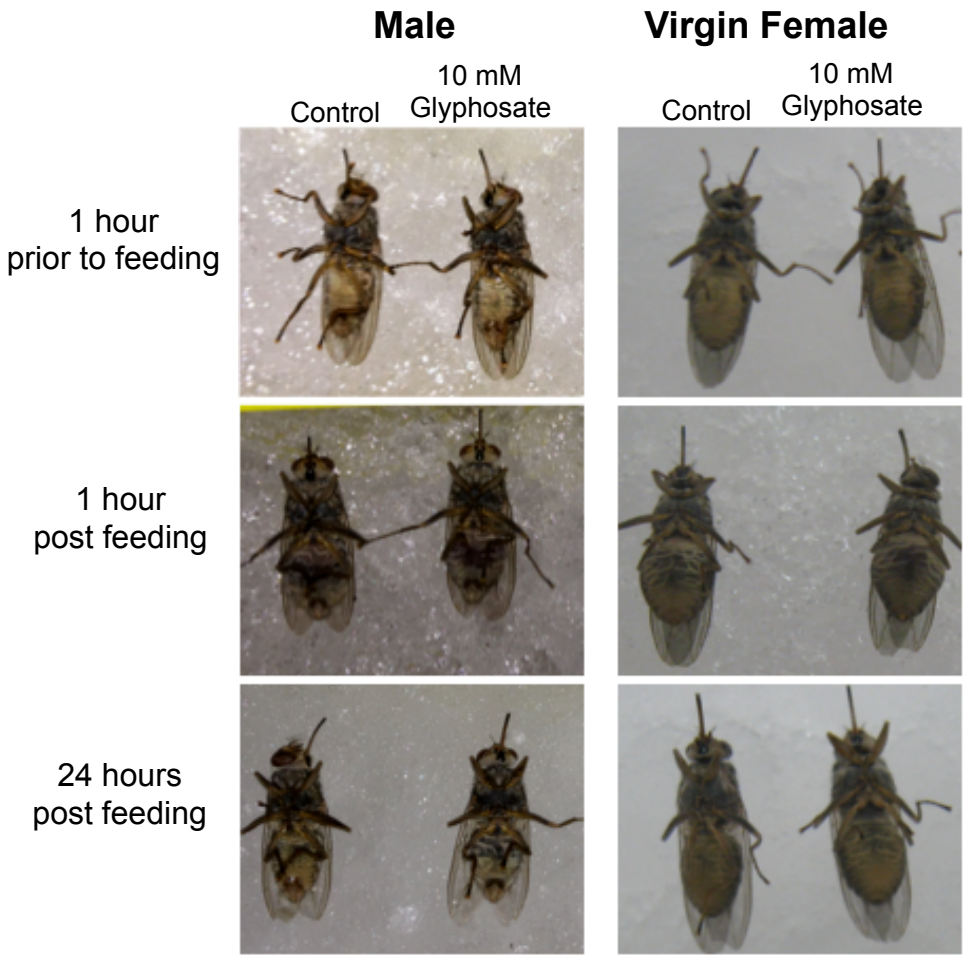


Figure 3-S4. Glyphosate administration decreases blood meal digestion in both sexes.

Representative pictures of fly abdomens at specified times pre/post feeding are depicted.

Table 3-1. Primers used in this study. qPCR primers used in this study were designed using Beacon Designer 7.2 software (Premier Biosoft International) and internal standard primers were created using Primer3 (http://frodo.wi.mit.edu/).

Primer name	Sequence (5' – 3')	Ta (°C)	Length (bp)
aroAqPCRFor	GGT ACT GCT ATG CGT TTG C	54.3	128
aroAqPCRRev	GCA CCA CCT TGT CTT AAA GC		
pabBqPCRFor	CGC AAA TTG GAA CCG TAT CAG	50.2	154
pabBqPCRRev	CCC GTA ATT GAC CCA CCT G		
folPqPCRFor	TTT CTG ATG GTG GAC AGT TTA C	51.5	149
folPqPCRRev	TCG TTC TGA TTC TTC AAG TTC G		
rpsCqPCRFor	CAA GAC CTG GAA TAG TAA TCG G	50.1	198
rpsCqPCRRev	CAC GCT TCA TTG CTC TAC G		
aroAstdFor	TTT TAT TAT CGG CGC AAA CC	55.0	457
aroAstdRev	AAT GGG GCC ATG ATG AGT AA		
pabBstdFor	TAA CTG CGC ACC ATT TTC TG	52.0	468
pabBstdRev	CCA CAC CAT GCA TTT CTT CT		
folPstdFor	AAA TGT CAC ACC GGA TTC GT	55.0	498
folPstdRev	CCG GGG TCA ATG ATT ATA CG		
rpsCstdFor	TGG CGT TCT ACA TGG TAT GC	55.0	702
rpsCstdRev	TGC ACG AAA AGT GTG TAG GG		
GmBtubfor	CCA TTC CCA CGT CTT CAC TT	55.0	151
GmBtubrev	GAC CAT GAC GTG GAT CAC AG		
GmRFCfor	CTC AAA GCC ACC ACC TTG TT	55.0	294
GmRFCrev	CAA CGA TGA CAA GAC GGC TA		

CHAPTER 4: The use of hypervariable genome regions to examine divergence of *Sodalis* and allied symbionts

CHAPTER 4.1: The phylogeny of *Sodalis*-like symbionts as reconstructed using surface-encoding loci*

ABSTRACT

Phylogenetic analyses of 16S rRNA support close relationships between the Gammaproteobacteria *Sodalis glossinidius*, a tsetse (Diptera: Glossinidae) symbiont, and bacteria infecting diverse insect orders. To further examine the evolutionary relationships of these *Sodalis*-like symbionts, phylogenetic trees were constructed for a subset of putative surface-encoding genes (i.e. *ompA*, *spr*, *slyB*, *rcsF*, *ycfM*, and *ompC*). The *ompA* and *ompC* loci were used toward examining the intra- and interspecific diversity of *Sodalis* within tsetse, respectively. Intraspecific analyses of *ompA* support elevated nonsynonymous (dN) polymorphism with an excess of singletons, indicating diversifying selection, specifically within the tsetse *Glossina morsitans*. Additionally, interspecific *ompC* comparisons between *Sodalis* and *Escherichia coli* demonstrate deviation from neutrality, with higher fixed dN observed at sites associated with extracellular loops. Surface-encoding genes varied in their phylogenetic resolution of *Sodalis* and related bacteria, suggesting conserved vs. host-specific roles. Moreover, *Sodalis* and its close relatives exhibit genetic divergence at the *rcsF*, *ompA*, and *ompC* loci, indicative of initial molecular divergence. The application of outer membrane genes as markers for further delineating the systematics of recently diverged bacteria is discussed. These

* Reprinted from Snyder AK, McMillen CM, Wallenhorst P, and Rio RVM. 2011. The phylogeny of *Sodalis*-like symbionts as reconstructed using surface encoding loci. *FEMS Microbiol. Lett.* 317: 143-151.

results increase our understanding of insect symbiont evolution, while also identifying early genome alterations occurring upon integration of microorganisms with eukaryotic hosts.

INTRODUCTION

Symbiosis enables the utilization of environments that would otherwise be rendered inhospitable and as such, is recognized as an important source of biological innovations particularly in regards to the radiation of the Class *Insecta* (Blochmann, 1887; Buchner, 1965). The evolutionary trajectory of symbiosis towards obligate mutualism may develop through a parasitism to mutualism continuum through processes such as the attenuation of host fitness penalties (Jeon, 1972) and the conversion of horizontal transmission to a purely vertical mode (Ewald, 1987). Such a route is exemplified by ancient endocellular symbionts of various insect hosts, such as *Buchnera aphidicola* in aphids (Homoptera: Aphididae), which are thought to have evolved from less specialized but more prevalent microbial relations such as those involving general insect pathogens (Dale *et al.*, 2001; Hosokawa *et al.*, 2010).

The gamma-proteobacterium, *Sodalis glossinidius*, is the secondary symbiont of the tsetse fly (Diptera: Glossinidae). Tsetse flies have medical significance as obligate vectors of the parasitic *Trypanosoma brucei* ssp., the etiological agents of African trypanosomiasis. In contrast to the primary symbiont *Wigglesworthia glossinidia*, which has a strict localization to the tsetse bacteriome and an extensive coevolutionary history with its host (Chen *et al.*, 1999), *Sodalis* exhibits a wider tissue tropism including the host midgut, hemolymph, and muscle (Cheng & Aksoy, 1999) with the symbiosis being of relatively recent origin (Weiss *et al.*, 2006). The functional role of *Sodalis* within tsetse remains relatively unknown, although influences on enhancing host life longevity (Dale & Welburn, 2001) and vector competency (Welburn *et al.*, 1993; Farikou *et al.*, 2010) have been demonstrated.

Recent studies have shown that symbionts harbored within several host insect orders including Diptera, Coleoptera, Phthiraptera, and Hemiptera are highly related to *Sodalis* based on 16S rRNA gene sequences (Weiss *et al.*, 2006; Fukatsu *et al.*, 2007; Novakova & Hyspa, 2007; Grunwald *et al.*, 2010; Kaiwa *et al.*, 2010; Toju *et al.*, 2010). These analyses indicate that this group of bacteria shares a recent common ancestor, despite now infecting a broad taxonomic range of hosts.

Selection pressures unique to ecological niches drive evolutionary diversification, with genomic alterations facilitating the adaptation to new habitats by bacteria. Outer membrane proteins, with known immunogenic properties, represent initial points of interspecific contact. Moreover, symbiont cell surfaces have been shown to be pivotal toward the homeostasis of host–bacterial relations (Weiss *et al.*, 2008; Nyholm *et al.*, 2009). Among related microorganisms, genes encoding surface-associated proteins are likely to represent preliminary examples of divergence due to host background differences and consequential symbiont adaptation. We believe that surface-encoding genes, often representing hypervariable genes (Wimley, 2003; Zheng *et al.*, 2003), may prove to be significant markers not only in deciphering the evolutionary distance between recently diverged microorganisms such as the *Sodalis*-allied bacteria, but also toward identifying preliminary molecular alterations associated with inhabiting diverse hosts.

For this study, we extend molecular phylogenetic analyses for this specific clade of *Sodalis*-like insect symbionts, particularly focusing on the symbionts of the tsetse fly species *Glossina morsitans*, *Glossina brevipalpis*, *Glossina fuscipes*, and *Glossina pallidipes*, the slender pigeon louse *Columbicola columbae* (Phthiraptera: Philopteridae), and the bloodsucking hippoboscid fly *Craterina melbae* (Diptera: Hipposboscidae). We aim to further our understanding of their relatedness and identify initial effects associated with the colonization of

different host species. The goals of the current study are: to assess intra/interspecies diversity of *Sodalis*, to provide 16S rRNA gene phylogenetic analysis of all '*Sodalis*-allied' microorganisms described to date, and to compare the ability of surface encoding genes to systematically resolve relationships within this symbiont lineage.

MATERIALS AND METHODS

Insects

Tsetse flies, *G. morsitans* and *G. brevipalpis*, were maintained at West Virginia University within the Department of Biology insectary as described previously (Snyder *et al.*, 2010).

Interspecific diversity analyses

DNA isolation (*C. melbae*, *G. morsitans*, *G. fuscipes*, *G. pallidipes*, and *G. brevipalpis*) was performed using the Holmes–Bonner protocol (Holmes & Bonner, 1973). Nucleic acid extraction for *C. columbae* was performed using the QIAamp tissue mini kit (Qiagen, Valencia, CA). All samples were resuspended in 1X Tris-EDTA following DNA isolation. DNA samples were subjected to PCR amplification of genes encoding putative outer membrane components; specifically *ompA*, the outer membrane protein A, *ompC*, the osmoporin protein C, and *rcsF*, *ycfM*, *slyB*, and *spr*, producing various outer membrane lipoproteins. PCR annealing temperatures, primers, and respective amplicon sizes are included in Supporting Information, Table S1. Notably, amplification reactions of *ycfM* from *C. columbae* and *C. melbae* and *rcsF* and *slyB* from *C. columbae* were not successful. Negative controls were included in each set of amplification reactions. The amplification products were analyzed by agarose gel electrophoresis and visualized with Kodak 1D image analysis software. The amplicons were purified using QIAquick PCR purification kit (Qiagen) and subject to DNA sequencing at the West Virginia University's Department of Biology Genomics Center on an ABI 3130xl analyzer (Applied

Biosystems, Foster City, CA) using a 3.1 BigDye protocol (Applied Biosystems). For each sample, three to five amplicons were sequenced in both directions and contigs were assembled using Ridom Trace Edit (Ridom GmbH, Wurzburg Germany).

Assessing *Sodalis* intraspecies diversity within tsetse

The *Sodalis ompA* gene was amplified from two *G. morsitans*, *G. fuscipes*, *G. brevipalpis*, and *G. pallidipes* individuals. Amplicons were ligated into pGEM-T vector (Promega) and *Escherichia coli* JM109 cells were transformed. Four colonies per individual tsetse were verified for an *ompA* insertion and sequenced as described above.

Molecular phylogenetic analyses

All analyses included sequence data collected in this study or publicly available at NCBI GenBank. DNA sequences were aligned using the CLUSTAL X algorithm with default settings, and refined manually when necessary. Maximum parsimony (MP) and neighbor joining (NJ) analyses were performed with 1000 replicates in PAUP 4.0 (Swofford, 2002). MP heuristic searches utilized the tree-bisection-reconnection (TBR) branch-swapping algorithm with 200 Max trees and starting trees were created using stepwise additions. All MP analyses were performed twice, where gaps were treated either as 'missing data' or as a 'fifth character state,' with no differences noted between the results. NJ analyses implemented Kimura's two-parameter model (Kimura, 1980). Lineage support was measured by calculating nonparametric bootstrap values (n = 1000) (Felsenstein, 1985).

The evolutionary models used for Bayesian analyses were determined using the Akaike Information Criterion in MRMODELTEST 2.3 (Nylander, 2004). Bayesian analyses were performed in MRBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003), and the number of categories used to approximate the gamma distribution was set at four. Additionally, six Markov chains

(Larget & Simon, 1999) were run for 3,000,000 generations for 16S rRNA gene and for 1,000,000 generations for surface-encoding genes. Posterior probability (PP) values were subsequently calculated. Stabilization of model parameters (burn-in) occurred around 2,400,000 and 800,000 generations for 16S rRNA and surface-encoding genes, respectively. Every 100th tree after stabilization (burn-in) was sampled to calculate a 50% majority-rule consensus tree. All trees were constructed using the program FIGTREE v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Genetic divergence analyses

DNASP (Librado & Rozas, 2009) was used to calculate synonymous (dS) and nonsynonymous (dN) rates and two common measures of nucleotide variati on, π and θ_W, for determining *ompA* intraspecies variation within *Glossina*. Neutrality tests were also performed in DNASP. The McDonald–Krietman test and neutrality index (NI) were calculated by comparing the ratio of dS to dN mutations within either individual *Glossina* species for *ompA*, or among *Glossina* isolates for *ompC*, and an *E. coli* outgroup. The outgroup was composed of ecologically diverse *E. coli* representatives NC_000913, NC_008253, and NC_002655. These adaptive evolution tests have been shown to be most powerful when taxa are closely related (Clark *et al.*, 2003). We chose *E. coli* as our representative outgroup because it is a close relative of *Sodalis*, and has a wide representation of publicly available genome strains.

Nucleotide sequence accession numbers

The nucleotide sequences determined in this study have been deposited in the NCBI GenBank database under accession numbers HM626140–HM626149 and HQ914651-HQ914697.

RESULTS AND DISCUSSION

Phylogenetic placement of tsetse fly secondary symbionts (*Sodalis*) based on 16S rRNA gene analyses

To examine the evolutionary relationships of the newly identified *Sodalis*-like symbionts, we constructed phylogenetic trees based on 16S rRNA gene sequences. Bayesian analysis supports the monophyly of Gammaproteobacteria symbionts isolated from diverse insect orders (i.e. Diptera, Coleoptera, Hemiptera, and Phthiraptera) (Fig. 1). In general, there is a tight clustering of symbionts with respective insect host Order. Our Bayesian analysis also suggests the closer relationship of hippoboscid symbionts to weevil and pigeon louse symbionts, rather than to *Sodalis*, despite a common ancestry of their respective hosts within the Hippoboscoidea (Petersen *et al.*, 2007), thus further substantiating a previous hypothesis of independent symbiont acquisition events by these hosts (Novakova & Hyspa, 2007). However, there is only moderate Bayesian support for this relationship (PP = 77, data not shown) that is further decreased (PP = 51) when symbionts of the recently reported chestnut weevil *Curculio sikkimensis* (Toju *et al.*, 2010) and the stinkbug *Cantao ocellatus* (Kaiwa *et al.*, 2010) are included in the analyses. Analyses were unable to resolve the relationships of the symbionts harbored within the hippoboscid, chestnut weevil, and stinkbug indicative of relatively recent establishments and inadequate time for 16S rRNA gene diversification, or alternatively the transfer of these symbionts within these insect orders. With Bayesian analysis, symbiont relationships within the *Sitophilus* clade are highly resolved in comparison with that of *Sodalis*, where the scattering of host species (i.e. not reflective of *Sitophilus* speciation; Conord *et al.*, 2008) suggests independent acquisition within species. It is possible that horizontal transmission, in addition to the previously described vertical route (Heddi *et al.*, 1999), may also contribute to this

phylogenetic patterning of symbionts; this warrants further study. Interestingly, although bacterial endosymbiosis is believed to be old within weevils (dating back approximately 125 Myr), symbiont replacement is believed to have occurred multiple times in *Sitophilus* weevils with causative factors remaining speculative (Conord *et al.*, 2008).

Sodalis isolated from *in vitro* culture maintained through serial passage formed its own monophyletic clade, supporting diversification from current *Glossina* isolates. While culture isolates were grouped together based on the 16S rRNA gene, *Sodalis* obtained from the same host species did not follow this pattern (i.e. symbionts within *G. fuscipes*, *G. austeni*, and *G. palpalis*) suggesting either no diversity between tsetse fly isolates or the lack of resolution due to the conserved nature of this locus. Distance analyses of the 16S rRNA gene also support the higher similarity of bacteria within the *Sodalis* clade, relative to that housing the *Sitophilus* symbionts (data not shown), which may explain why analyses were unable to further resolve these relations (Fig. 1). Importantly, many branches could not be robustly resolved warranting the need for additional inquiries utilizing genes that are typically associated with higher evolutionary rates such as those encoding surface-exposed molecules.

Phylogenetic placement of *Sodalis*–like symbionts based on surface-encoding proteins

To further our understanding of the divergence of '*Sodalis*-allied' bacteria, particularly those found within various *Glossina spp*., *C. columbae*, and *C. melbae*, and to also assess the application of these surface encoding genes in future analyses extending into other related symbionts, we reconstructed their phylogeny using six putative outer membrane- encoding genes: *rcsF*, *slyB*, *ompA*, *spr*, *ompC*, and *ycfM*. With only a few exceptions (all *spr* and *Glossina* vs. *C. melbae slyB* comparisons), the genetic distances of surface-encoding loci between symbionts localized within hosts of different orders were greater in comparison with 16S rRNA

gene.

In regards to the *spr*, *slyB*, and *ycfM* loci, although sufficient sequence similarities resulted in the *Sodalis*-like isolates forming a monophyletic clade within the Gammaproteobacteria distinct from many free-living members of this group, deeper taxonomic resolution was lacking (data not shown). The low phylogenetic signal provided by these loci suggests that they may not be involved in adapting to particular host species and/or may be structurally constrained. For example, comparative analyses of the *spr* lipoprotein amino acid sequence demonstrated the conservation of residues that form a unique Cys–His–His catalytic triad that is believed to form a substrate-binding cleft within the active site of this protein (Aramini *et al.*, 2008) between examined *Sodalis* isolates, *C. melbae*, and *C. columbae* symbionts.

The *ompA*, *ompC*, and *rcsF* loci (Fig. 2) appear to be more informative toward the phylogenetic resolution of the *Sodalis*-like symbiont clade. With *rcsF*, sufficient phylogenetic signal was provided to enable clustering of the *Glossina* symbionts, with strong support, separate from the *C. melbae* symbiont (Fig. 2b). Interestingly, *rcsF* in *E. coli* has been shown to be involved in signaling transduction of perturbations and/or environmental cues from the cell surface (Majdalani *et al.*, 2005). Diversification between *Sodalis* and *C. melbae* isolates may indicate functional adaptations, such as differences in the type of signaling encountered within the host species background. The *Sodalis* symbionts also formed a distinct clade with the *ompC* phylogeny, with most mutations noted outside of the seven putative extracellular loops (Basle *et al.*, 2006) of the different *Glossina* isolates. The one exception occurred in extracellular loop 4, where host interspecies diversity was observed with *Sodalis* isolates.

Relative to the other surface encoding genes analyzed in this study, the *ompA* gene

exhibited the greatest diversity among symbionts due to a combination of point mutations and indels. The best-studied *ompA* gene variant, that of *E. coli* K-12, encodes a 325 amino acid polypeptide (Chen *et al.*, 1980). The N-terminal domain forms an eight-stranded β-barrel in the outer membrane, creating four surface-exposed loops (Pautsch & Schulz, 1998), while the C-terminus is periplasmic (Klose *et al.*, 1988). Amino acid variations within outer membrane proteins mainly occur in the domains located in the extracellular regions, while interspaced residues making up the β-strands tend to be conserved. In our analyses, relative to *Glossina* symbionts, a total of nine nonsynonymous mutations were observed among *C. melbae*, *C. columbae*, and *Sitophilus* (i.e. *Sitophilus oryzae* primary symbiont, SOPE) symbionts occurring in loops 1–4 of the OmpA protein. Differences noted in the *ompA* sequence between the *Glossina* symbionts were localized outside of the extracellular regions, similar to our observations with *ompC*. In relation to *ompA*, the *C. columbae* symbiont exhibited the greatest nucleotide divergence resulting in its sister taxon placement relative to the other symbionts of interest with strong MP bootstrap support. MP, Bayesian, and NJ analyses all grouped *Glossina* symbionts within their own clade indicative of diversification potentially arising from host adaptation processes.

Molecular evolution of *Sodalis* -like symbionts

The *Sodalis ompA* gene demonstrated a wide nucleotide variation (π) within tsetse species (Table 1), with the highest π exhibited within *G. morsitans* ($\pi = 0.11$) and the lowest within G. brevipalpis ($\pi = 0.001$). This observation is not unprecedented as evidence of endosymbiont genomes (e.g. *Wolbachia*) undergoing either purifying or diversifying selection when examined from different host species has also been described with cell envelope component genes (Brownlie *et al.*, 2007).

Tests of neutrality (Tajima's *D*, Fu and Li's *D** and *F**, and Fu and Li's *D* and *F*) indicate a significant excess of young, rare alleles for *Sodalis ompA* within *G. morsitans* and *G. pallidipes*. In summation, three indices (π, *dN/dS*, and NI) support diversifying selection due to an abundance of low frequency *Sodalis ompA* haplotypes within *G. morsitans*. These observations may reflect the well-supported phenomenon of enhanced sequence evolution in endosymbiotic bacteria (Clark *et al.*, 1999; Canback *et al.*, 2004; Fry & Wernegreen, 2005). Similar to other endosymbionts, the small effective population size of *Sodalis*, a consequence of severe population bottlenecks during maternal transmission (Rio *et al.*, 2006), predicts a larger proportion of nonsynonymous mutations due to drift that will generate higher *dN* to *dS* ratios (Ohta, 1972; Woolfit & Bronham, 2003).

Deviation from neutrality was also observed with *Sodalis ompC* isolates, as supported by a significant MK test (G= 13.42, $P = 0.00025$) when compared with *E. coli*. A high abundance of fixed *dN* substitutions within all *Sodalis* isolates provides strong evidence for positive selection at particular sites of the *ompC* gene. Notably, upon comparison of *Sodalis* with *E. coli* isolates, greater *ompC* amino acid sequence variation was observed at putative surface-exposed loops suggesting their significance in adaptive evolution toward ecological niches.

Here, we describe early genetic modifications likely involved in host adaptation within *Sodalis*-allied bacteria, specifically divergence in symbiont surface-encoding genes. In general, this particular class of loci exhibited greater genetic distances among *Sodalis*-like bacteria than the 16S rRNA gene traditionally used in phylogenetic analyses. Nevertheless, not all the surface-encoding genes examined in this study proved equivalent in their ability to resolve phylogenetic relations. Differences in selective pressures arising from distinct host physiologies and feeding lifestyles (Rio *et al.*, 2003; Toh *et al.*, 2006), as well as the influence of other host microbiota

members (Snyder *et al.*, 2010) have been shown to affect symbiont genome evolution. Future studies should extend the phylogenetics of these surface-encoding loci, specifically *rcsF*, *ompC*, and *ompA*, to other recently identified *Sodalis*-related symbionts to enhance phylogenetic resolution. Functional assays should be pursued also to examine the relevance of surface-encoding loci toward the process of endosymbiotic adaptation and to determine whether the described differences are sufficient to constrict host species colonization.

REFERENCES

1. Aramini JM, Rossi P, Huang YJ *et al.* (2008) Solution NMR structure of the NlpC/P60 domain of lipoprotein Spr from *Escherichia coli*: structural evidence for a novel cysteine peptidase catalytic triad. *Biochemistry* **47**: 9715–9717.
2. Basle A, Rummel G, Storici P, Rosenbusch JP & Schirmer T (2006) Crystal structure of osmoporin OmpC from *E. coli* at 2.0 A. *J Mol Biol* **362**: 933–942.
3. Blochmann F (1887) Uber das regel massige Vorkommen von backterienahnlichen Geblin in den Geweben und Eiern vershiedemer Insecten. *Z Biol* **24**: 1–16.
4. Brownlie JC, Adamski M, Slatko B & McGraw EA (2007) Diversifying selection and host adaptation in two endosymbiont genomes. *BMC Evol Biol* **7**: 68.
5. Buchner P (1965) *Endosymbiosis of Animals with Plant Microorganisms*. John Wiley & Sons, New York.
6. Canback B, Tamas I & Andersson SG (2004) A phylogenomic study of endosymbiotic bacteria. *Mol Biol Evol* **21**: 1110–1122.
7. Chen R, Schmidmayr W, Kramer C, Chen-Schmeisser U & Henning U (1980) Primary structure of major outer membrane II (OmpA protein) of *Escherichia coli* K-12. *P Natl Acad Sci USA* **77**: 4592–4596.
8. Chen X, Li S & Aksoy S (1999) 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.
9. Cheng Q & Aksoy S (1999) Tissue tropism, transmission, and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. *Insect Mol Biol* **8**: 125–132.
10. Clark AG, Glanowski S, Nielsen R *et al.* (2003) Inferring non-neutral evolution from

human–chimp–mouse orthologous gene trios. *Science* **302**: 1960–1963.

11. Clark MA, Moran NA & Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol Biol Evol* **16**: 1586–1598.
12. Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G & Heddi A (2008) Long-term evolutionary stability of bacterial endosymbiosis in Curculionidae: additional evidence of symbiont replacement in the Dryopthoridae family. *Mol Biol Evol* **25**: 859–868.
13. Dale C & Welburn SC (2001) The endosymbionts of tsetse flies – manipulating host–parasite interactions. *Int J Parasitol* **31**: 628–631.
14. Dale C, Young SA, Haydon DT & Welburn SC (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *P Natl Acad Sci USA* **98**: 1883–1888.
15. Ewald P (1987) Transmission modes and the evolution of the parasitism–mutualism continuum. *Ann NY Acad Sci* **503**: 295–306.
16. Farikou O, Njiokou F, Mdiba Mdiba JA, Njitchouang GR, Djeunga HN, Asonganyi T, Simarro PP, Cuny G & Geiger A (2010) 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.
17. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
18. Fry AJ & Wernegreen JJ (2005) The roles of positive and negative selection in the molecular evolution of insect endosymbionts. *Gene* **355**: 1–10.
19. Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, Sasaki- Fakatus K, Yoshizawa K,

Dale C & Clayton DH (2007) Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. *Appl Environ Microb* **73**: 6660–6668.

20. Grunwald S, Pilhofer M & Holl W (2010) Microbial associations in guts systems of wood- and bark-inhabiting long horned beetles [Coleoptera: Cerambyicidae]. *Syst Appl Microbiol* **33**: 25-34.
21. Heddi A, Grenier A, Khatchadourian C, Charles H & Nardon P (1999) Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *P Natl Acad Sci USA* **96**: 6814–6819.
22. Holmes DS & Bonner J (1973) Preparation, molecular weight, base composition, and secondary structure of giant ribonucleic acid. *Biochemistry* **12**: 2330–2338.
23. Hosokawa T, Koga R, Kikuchi Y, Meng XY & Fukatsu T (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Appl Environ Microb* **107**: 769–774.
24. Jeon K (1972) Development of cellular dependence on infective organism: microsurgical studies on amoebas. *Science* **179**: 1122–1123.
25. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M & Fukatsu T (2010) Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the scutellerid stinkbug *Cantao ocellatus*. *Appl Environ Microb* **76**: 3486–3494.
26. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequence. *J Mol Evol* **16**: 111–120.
27. Klose M, Schwarz H, MacIntyre S, Freudl R, Eschbach ML & Henning U (1988) Internal deletions in the gene for an *Escherichia coli* outer membrane protein define an area possibly important for recognition of the outer membrane of this polypeptide. *J Biol*

Chem **263**: 13291–13296.

28. Larget B & Simon DL (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol Biol Evol* 16: 750–759.
29. Librado P & Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
30. Majdalani N, Heck M, Stout V & Gottesman S (2005) Role of RcsF in signaling to the Rcs phosphorelay pathway in *Escherichia coli. J Bacteriol* **187**: 6770–6778.
31. Novakova E & Hyspa V (2007) A new *Sodalis* lineage from the bloodsucking fly *Craterina melbae* (Diptera: Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS Microbiol Lett* **269**: 131–135.
32. Nyholm SV, Stewart JJ, Ruby EG & McFall-Ngai MJ (2009) Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environ Microbiol* **11**: 483–493.
33. Nylander JAA (2004) *MrModeltest v2. Evolutionary Biology Centre*. Uppsala University, Uppsala, Sweden.
34. Ohta T (1972) Population size and the rate of evolution. *J Mol Evol* **1**: 305–314.
35. Pautsch A & Schulz GE (1998) Structure of the outer membrane protein A transmembrane domain. *Nat Struct Biol* **5**: 1013–1017.
36. Petersen FP, Meier R, Narayananan Kutty S & Wiegmann BM (2007) The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. *Mol Phylogenet Evol* **45**: 111–122.
37. Posada D & Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

38. Rio RV, Lefevre C, Heddi A & Aksoy S (2003) Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. *Appl Environ Microb* **69**: 6825–6832.

39. Rio RVM, Wu Y, Filardo G & Aksoy S (2006) Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proc Biol Sci* **273**: 805–814.

40. Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

41. Snyder AK, Deberry JW, Runyen-Janecky L & Rio RVM (2010) Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc Biol Sci* **277**: 2389–2397.

42. Swofford DL (2002) *PAUP 4.0-Phylogenetic Analysis Using Parsimony. Version 4.* Sinauer Associates, Sunderland, MA.

43. Thompson JD, Higgins DG & Gibson JJ (1994) Clustal W: improving the sensitivity of progressive multiple alignments through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

44. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M & Aksoy S (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* **16**: 149–156.

45. Toju H, Koga R, Nikoh N, Meng XY, Kimura N & Fukatsu T (2010) ‘‘Candidatus *Curculioniphilus buchneri*’’ a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. *Appl Environ Microb* **76**: 275–282.

46. Weiss BL, Mouchotte RM, Rio RVM, Wu Y, Wu Z, Heddi A & Aksoy S (2006)

Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. *Appl Environ Microb* **72**: 7013–7021.

47. Weiss BL, Wu Y, Schwank JJ, Tolwinski NS & Aksoy S (2008) An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. *P Natl Acad Sci USA* **105**: 15088–15093.
48. Welburn SC, Arnold K, Maudlin I & Goday GW (1993) *Rickettsia*-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology* **107**: 141–145.Wimley WC (2003) The versatile beta-barrel membrane protein. *Curr Opin Struc Biol* **13**: 404–411.
49. Woolfit M & Bronham L (2003) Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol Biol Evol* **20**: 1545–1555.
50. Zheng Y, Roberts RJ & Kasif S (2003) Identification of genes with fast-evolving regions in microbial genomes. *Nucleic Acids Res* **278**: 24825–24830.

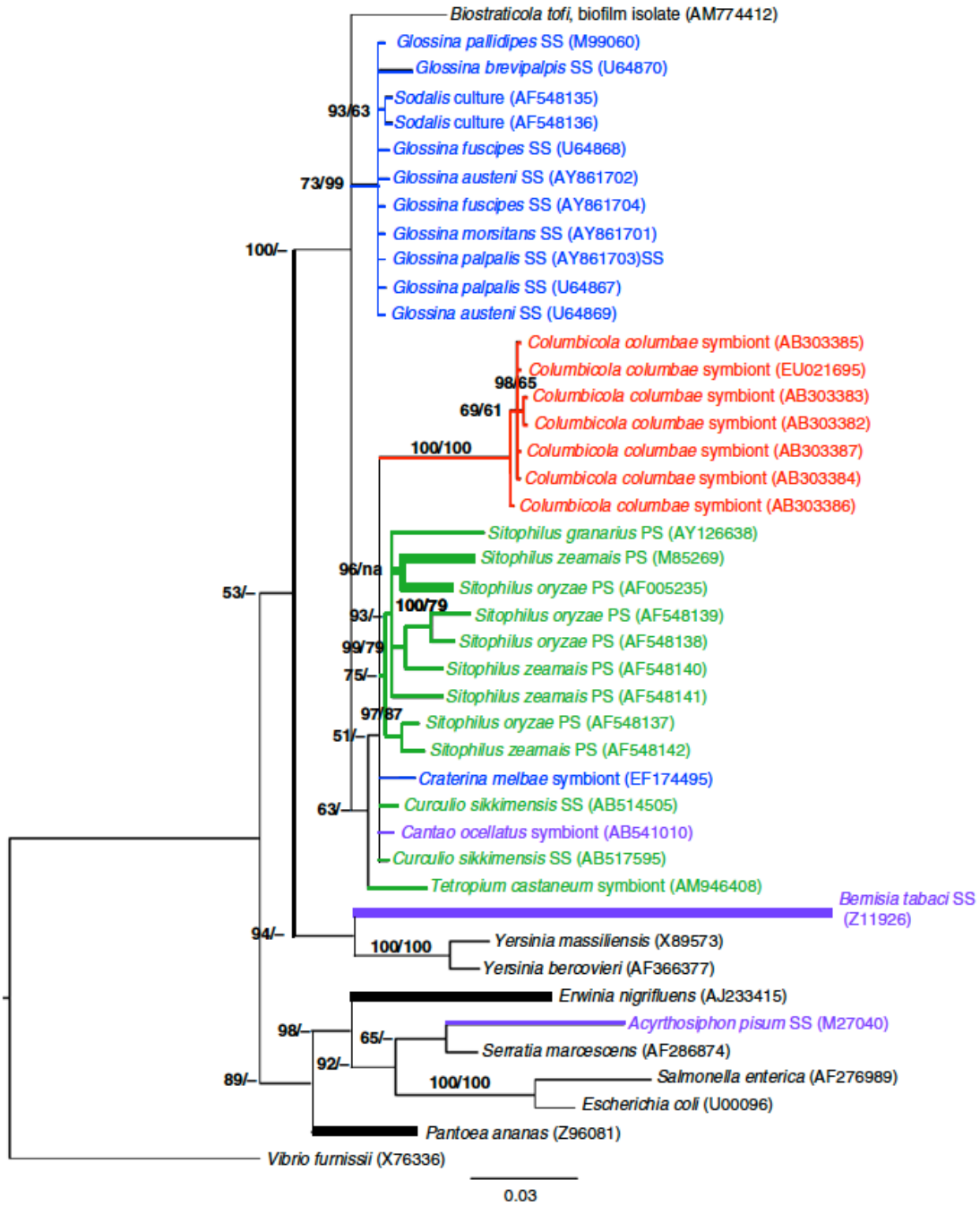

Biostraticola tofi, biofilm isolate (AM774412)
Glossina pallidipes SS (M99060)
Glossina brevipalpis SS (U64870)
93/63
Sodalis culture (AF548135)
Sodalis culture (AF548136)
Glossina fuscipes SS (U64868)
73/99
Glossina austeni SS (AY861702)
Glossina fuscipes SS (AY861704)
Glossina morsitans SS (AY861701)
100/–
Glossina palpalis SS (AY861703)SS
Glossina palpalis SS (U64867)
Glossina austeni SS (U64869)
Columbicola columbae symbiont (AB303385)
Columbicola columbae symbiont (EU021695)
98/65
Columbicola columbae symbiont (AB303383)
69/61
Columbicola columbae symbiont (AB303382)
100/100
Columbicola columbae symbiont (AB303387)
Columbicola columbae symbiont (AB303384)
Columbicola columbae symbiont (AB303386)
Sitophilus granarius PS (AY126638)
Sitophilus zeamais PS (M85269)
96/na
Sitophilus oryzae PS (AF005235)
53/–
100/79
93/–
Sitophilus oryzae PS (AF548139)
Sitophilus oryzae PS (AF548138)
99/79
Sitophilus zeamais PS (AF548140)
75/–
Sitophilus zeamais PS (AF548141)
97/87
Sitophilus oryzae PS (AF548137)
51/–
Sitophilus zeamais PS (AF548142)
Craterina melbae symbiont (EF174495)
63/–
Curculio sikkimensis SS (AB514505)
Cantao ocellatus symbiont (AB541010)
Curculio sikkimensis SS (AB517595)
Tetropium castaneum symbiont (AM946408)
Bemisia tabaci SS (Z11926)
94/–
100/100
Yersinia massiliensis (X89573)
Yersinia bercovieri (AF366377)
Erwinia nigrifluens (AJ233415)
98/–
65/–
Acyrthosiphon pisum SS (M27040)
Serratia marcescens (AF286874)
92/–
89/–
100/100
Salmonella enterica (AF276989)
Escherichia coli (U00096)
Pantoea ananas (Z96081)
Vibrio furnissii (X76336)
0.03

Figure 4.1-1. Molecular phylogenetic tree of 16S rRNA gene sequences from *Sodalis* and allied bacteria. A Bayesian analysis tree created from 1509 aligned nucleotides is shown; NJ analyses gave essentially identical results (data not shown). Branches in bold were constrained with MP analysis. PP (shown as %, i.e. 95% represents a PP value of 0.95) and bootstrap values > 50% are indicated at the nodes (-, < 50% bootstrap), respectively. The branch lengths are measured in expected substitutions per site. Sequence accession numbers are provided. Host species are indicated for symbiotic bacteria, with colors representing insect orders. PS, primary symbiont; SS, secondary symbiont.

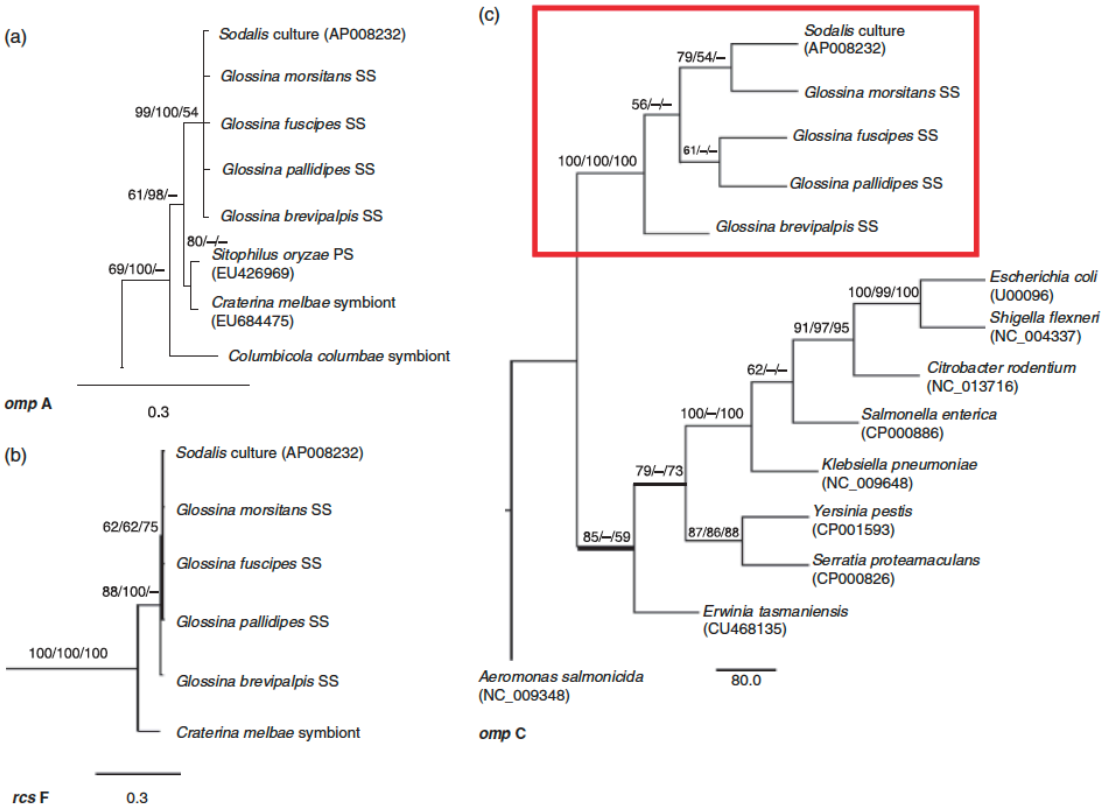


Figure 4.1-2. Molecular phylogenetic analyses of putative outer membrane encoding gene sequences from *Sodalis*-allied symbionts which support diversification. Bayesian trees inferred from (a) 1164 unambiguously aligned nucleotides of the *ompA* gene and (b) 426 nucleotides of the *rcsF* gene. Significance values are indicated in Bayesian PP/MP bootstrap /NJ bootstrap. Branch lengths are measured in expected substitutions per site and depicted under each tree. (c) MP tree inferred from 1227 nucleotides of the *ompC* genes are shown with support values in the order of MP bootstrap / Bayesian PP/NJ bootstrap. Branch lengths depict the number of substitutions. Bold lines indicate discrepancies in tree renditions between analyses. Accession numbers are provided in parentheses. Host species are indicated for symbiotic bacteria; SS, secondary symbiont.

Table 4.1-1. *Sodalis ompA* nucleotide diversity within tsetse species and tests for neutral models of evolution.

Symbiont host species	π_{Total}	θ_w	ω (*dN*/*dS*)	NI	Tajimas's *D*	Fu and Li's *D**	Fu and Li's *F**	Fu and Li's *D**	Fu and Li's *F*
Glossina morsitans	0.112	0.161	2.03	4.331***	− 1.75***	− 1.86**	− 2.04**	− 2.39**	− 2.72**
Glossina pallidipes	0.046	0.068	0.828	0.627	− 1.83***	− 1.97**	− 2.16**	− 2.19*	− 2.56**
Glossina fuscipes	0.004	0.006	0	1.000	− 1.13	− 1.3	− 1.37	− 1.49	− 1.65
Glossina brevipalpis	0.001	0.001	0	0.311	− 1.05	− 1.13	− 1.20	− 1.26	− 1.41

The neutrality index (NI), the ratio of synonymous to nonsynonymous mutations (*dN*/*dS*), was calculated using the McDonald–Kreitman test. Neutrality was examined within tsetse isolates (Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F**) and also compared with the outgroup *Escherichia coli* (accession number NC_000913) using Fu and Li's *D* and Fu and Li's *F*.

Statistical significance:

*$P < 0.05$; **$P < 0.02$; ***$P < 0.01$.

π, average pairwise nucleotide diversity; θ_w, segregating sites per haploid genome.

Table 4.1-S1. Primers, annealing temperatures (T_a), and resulting amplicon sizes.

Gene	Gene product	Primer	Primer sequence (5'-3')	T_a (°C)	Amplicon size (bp)*
*omp*A	Outer membrane protein A	F	acagctatcgcacttgcagt	55.0	991
		R	cggcctttcacgctgttaca		
spr	Outer membrane lipoprotein	F	atatgttctgcggctcatcc	55.0	515
		R	cttcacggtaacgggatttc		
*sly*B	Outer membrane lipoprotein	F	tgaaacgtttgatcgtggtg	55.0	459
		R	gcggggatacggtgatagt		
*rcs*F	Outer membrane lipoprotein	F	cctgtttcgctgttccttct	55.0	373
		R	ggcagatggcctgttgatag		
*omp*C	Outer membrane porin protein C	F	cggcatgcgctatatgtcta	55.0	964
		R	gtcacgggtgaagtcgtttt		
*ycf*M	Outer membrane lipoprotein	F	caagctgtacctcccgaaag	55.0	529
		R	ccgttaccggaccagataat		

* Relative to expected amplicon size of *Sodalis glossinidius* from '*in vitro*' culture.

CHAPTER 4.2: Use of the internal transcribed spacer (ITS) regions to examine symbiont divergence and as a diagnostic tool for *Sodalis*-related bacteria[*]

ABSTRACT

Bacteria excel in most ecological niches, including insect symbioses. A cluster of bacterial symbionts, established within a broad range of insects, share high 16S rRNA similarities with the secondary symbiont of the tsetse fly (Diptera: Glossinidae), *Sodalis glossinidius*. Although 16S rRNA has proven informative towards characterization of this clade, the gene is insufficient for examining recent divergence due to selective constraints. Here, we assess the application of the internal transcribed spacer (ITS) regions, specifically the ITS^{glu} and $ITS^{ala,ile}$, used in conjunction with 16S rRNA to enhance the phylogenetic resolution of *Sodalis*-allied bacteria. The 16S rRNA + ITS regions of *Sodalis* and allied bacteria demonstrated significant divergence and were robust towards phylogenetic resolution. A monophyletic clade of *Sodalis* isolates from tsetse species, distinct from other Enterobacteriaceae, was consistently observed suggesting diversification due to host adaptation. In contrast, the phylogenetic distribution of symbionts isolated from hippoboscid flies and various Hemiptera and Coleoptera were intertwined suggesting either horizontal transfer or a recent establishment from an environmental source. Lineage splitting of *Sodalis*-allied bacteria into symbiotic and free-living sister groups was also observed. Additionally, we propose an ITS region as a diagnostic marker for the identification of additional *Sodalis*-allied symbionts in the field. These results expand our knowledge of informative genome regions to assess genetic divergence since splitting from the last common ancestor, of this versatile insect

[*] Reprinted from Snyder AK, Adkins KZ and Rio RVM. 2011. Use of the internal transcribed spacer (ITS) regions to examine symbiont divergence and as a diagnostic tool for *Sodalis*-related bacteria. *Insects*. 2: 515-531.

symbiont clade that have become increasingly recognized as valuable towards our understanding of the evolution of symbiosis. These facultative and recently associated symbionts may provide a novel source of traits adaptable to the dynamic ecologies encountered by diverse host backgrounds.

INTRODUCTION

Symbioses abound in the class Insecta, where an extraordinary range of host effects, temporal and spatial distribution, and degree of co-evolution has been reported [1]. Symbioses are recognized as a widespread source of evolutionary innovation for insects. For example, insects whose diets are nutritionally unbalanced typically harbor symbionts referred to as primary symbionts (P-symbionts) that can provision essential metabolic supplementation [2], often enabling host niche expansion. P-symbiont establishment is assured through strict vertical transmission, thereby ensuring persistence of the relationship and resulting in lengthy co-evolution with its host [3,4].

Insects may also harbor facultative microbes known as secondary symbionts (S-symbionts). S-symbionts, although not obligate to host biology, may provide host benefits depending on environmental context, such as during periods of heat stress [5], parasitoid attack [6], or towards the utilization of particular host plant substrates [7]. Moreover, distantly related insects can harbor closely related bacterial S-symbionts, suggesting initial widespread microbial infection, most likely through horizontal transfer or as a free-living generalist with multiple independent host acquisitions [8]. Symbiotic establishment may then be followed by genomic tailoring through evolutionary time, leading to functional specialization complementary to host biology and ecology, similar to what has been reported with P-symbionts [9]. S-symbionts may accordingly represent intermediates in the evolutionary trajectory to an exclusively symbiotic

lifestyle [10].

Tsetse flies provide ideal biological models to examine symbiosis due to the presence of a low complexity microbiota, yet representing a wide range of host-microbe relations [11]. The tsetse microbiota predominantly consists of two Gammaproteobacteria; an obligate P-symbiont *Wigglesworthia glossinidia*, and a S-symbiont *Sodalis glossinidius*, as well an Alphaproteobacteria, the facultative parasite *Wolbachia pipientis* [12]. Additionally, tsetse flies maintain significant medicinal and socioeconomic importance as the vectors of African trypanosomiasis. Consequently, symbiotic microbes are also of applied interest as their genetic manipulation offers potential disease control mechanisms [13].

In contrast to the P-symbiont *Wigglesworthia*, *Sodalis* [14] has only recently associated with the tsetse host [15]. Evidence of a recent transition into symbiosis includes its wide host tissue tropism [16], amenability towards *in vitro* culture [17], stochastic presence within tsetse field populations [18,19], and a lack of congruence with tsetse phylogeny [20]. Genomic features [21], notably; a relatively larger ~4.2 Mb size, lack of A-T (Adenine-Thymine) bias, and the presence of phage-like and symbiosis region genes also support a recent transition into symbiosis. Despite these features, there are also some indications of *Sodalis* evolving into an endosymbiotic lifestyle such as a high proportion of pseudogenes with homologs of proteins involved in defense or in the transport and metabolism of carbohydrates and inorganic ions [21], believed to be unessential within the host. Furthermore, metabolic interplay resulting from genomic complementation between *Wigglesworthia* and *Sodalis* demonstrates early functional convergence, which may act to evade species antagonism [22].

Culture independent sequencing techniques have enabled the identification of numerous bacterial species residing within a diverse range of hosts, particularly insects [23]. One such

group gaining recognition, based on high 16S rRNA gene identity, comprises *Sodalis* and related bacteria within a broad range of insects, including various Hemiptera, Diptera, Coleoptera and a Phithiraptera [24-29]. Although the 16S rRNA gene has proven quintessential in many microbial phylogenetic studies, it's exclusive use is poorly suited to differentiate recently diverged bacteria (*i.e.*, genus level and below) due to the conserved regions lacking informative characters [28,30] and the potential occurrence of homoplasy or intraspecific variation within the hypervariable regions [31].

The conserved nature of the 16S rRNA locus has lead to the use of other genome regions for the phylogenetic analyses of closely related organisms. A recent application of outer membrane genes as markers for delineating the systematics of the *Sodalis* clade demonstrated sequence variation, notably in putative surface exposed loops, likely arising from the adaptive evolution towards particular host features, such as immunity [32]. An additional example, the internal transcribed spacer regions have been shown to exhibit an accelerated evolutionary rate relative to the conventionally used 16S rRNA gene [33]. Noncoding ITS regions that separate the 16S rRNA-23S rRNA and the 23S rRNA-5S rRNA are designated as ITS1 and ITS2, respectively. Additionally, the ITS regions may encode tRNAs. Use of the ITS regions have proven informative in both sequence and length variation for the phylogenetic resolution of bacterial species [34] and strains [35].

The molecular phylogenetic analyses of bacteria, from both free-living and host-associated lifestyles, may enhance our understanding of how environmental generalists transition into symbioses that become so specialized that they rely purely on vertical transmission and are associated with the evolution of extreme genome features. In this study, we have coupled the 16S rRNA and ITS regions to examine the phylogeny and diversity of *Sodalis*-allied symbionts

widely distributed throughout the class Insecta. Our results provide information on additional genetic variation among the *Sodalis*-like symbionts, further evidence to support diversification of this clade from an environmental progenitor and high likelihood for the lateral transfer of symbionts between diverse insect orders. Furthermore, we propose the ITS regions, used in conjunction with 16S rRNA, as a diagnostic tool for the identification and characterization of additional *Sodalis*-allied symbionts from insect hosts in the field. These symbionts provide snapshots of early events associated with the transitioning into insect symbiosis, and are potentially useful towards revealing both universal aspects of partner association as well as unique attributes towards particular symbioses. Methods that enhance our ability to detect these symbionts may increase the number of symbioses available for crucial comparative studies.

MATERIALS AND METHODS

Specimens and DNA Isolation

DNA was isolated from tsetse adult flies (*Glossina brevipalpis*, *G. morsitans*, *G. fuscipes*, and *G. pallidipes*), hippoboscid adult flies and pupae (*Craterina melbae*), larval stage chestnut weevils (*Curculio sikkimensis*) and *Sodalis* bacteria from *in vitro* culture following the Holmes-Bonner protocol [36]. Due to the sympatric localization of *Cu. sikkimensis* with the sister species *Cu. dentipes*, as well as the lack of distinguishable morphological features between the two species as larvae, the species identification was verified by sequencing of the mitochondrial cytochrome oxidase subunit I, *CO1* [37]. DNA samples of the ovaries of adult shieldbugs (*Eucorysses grandis*) and scutellerid stinkbugs (*Cantao ocellatus*) were obtained by using a NucleoSpin Tissue kit (Macherey-Nagel, Bethleham, PA). Additionally, the *Sodalis*-like *Biostraticola tofi* DNA, originally isolated from the biofilm of a tufa (porous rock formed by the precipitation of H_2O) deposit [38], was obtained from DSMZ (Braunschweig, Germany). All

samples were re-suspended in 1× Tris-EDTA following DNA isolation.

PCR Amplification and Sequencing of ITS Regions

To amplify the ITS1 regions, primers were designed to the 3' region of the *Sodalis* 16S rRNA gene (NC_007712; ITSfor: 5'-GGA GTG GGT TGC AAA AGA AG-3') and the 5' region of the 23S rRNA gene (ITSrev: 5'-CCA CCG TGT ACG CTT AGT CA-3') (Figure S1) using the default Primer3 algorithm [39]. DNA samples were subjected to PCR amplification in 50 μL reactions consisting of 1.25 U GoTaq Flexi DNA Polymerase (Promega, Madison, WI, USA), 4 mM $MgCl_2$, 1× Green GoTaq Flexi Buffer, and 0.2 mM dNTPs and primers. Amplification conditions consisted of 3 min initial denaturation at 95 °C, followed by 34 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1.5 min, with a final elongation at 72 °C for 10 min. Negative controls were included in all reactions.

The amplification products were analyzed by agarose gel electrophoresis and viewed using Kodak 1D image analysis software. Resulting amplicons of 600–1,000 bp were extracted and purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). Following gel extraction, amplicons were either sequenced or ligated into pGEM-T vector (Promega, Madison, WI, USA) and transformed using *Escherichia coli* JM109 cells (Promega).

Amplicons were sequenced at the West Virginia University Department of Biology Genomics Facility with an ABI 3130 × 1 analyzer (Applied Biosystems, Foster City, CA, USA) using a 3.1 BigDye protocol (Applied Biosystems). For each DNA sample, at least three amplicons were sequenced using both forward and reverse primers and contigs were assembled using Ridom Trace Edit (RidomGmbH, Wurzburg, Germany). If any nucleotide variation was observed, 5 additional clones were subsequently sequenced to assess ITS variation.

Molecular Phylogenetics

Consensus sequences were created from the contigs and edited to remove the 23S rRNA regions, so that only the 16S rRNA and ITS regions were analyzed. Sequences were aligned using MUSCLE [40] and inspected and corrected manually. Percent nucleotide identity between sequences was determined using PAUP 4.0 by comparing pairwise base differences [41].

Molecular phylogenetic analyses included Neighbor joining (NJ), Maximum parsimony (MP), and Bayesian methods. NJ and MP analyses were performed using PAUP 4.0 with the Kimura's two-parameter model of nucleotide substitution and 1,000 nonparametric bootstrap (BS) replicates, as a measure of lineage support. MP heuristic searches implemented 1,000 replicates using the tree-bisection-reconnection algorithm, where starting trees for branch swapping were obtained through random Stepwise-Additions, and Max trees set at 200. Additionally, each MP analysis was performed twice, with gaps treated as either "missing data" or "5th character state", with no differences noted among the resulting phylogenies.

Bayesian analyses were performed using MrBayes 3.1.2 [42] with Posterior Probabilities (PP) calculated. Evolutionary models to implement for each dataset where chosen using the Akaike Information Criterion in MrModelTest version 2.3 [43]. The best fit model implemented in both the 16S rRNA and ITS^{glu} or $ITS^{ala,ile}$ analyses was the General Time Reversible + invariant sites + gamma (GTR + I + G). Additionally, Markov chain Monte Carlo parameters were set to 6 chains and 1 million generations. Stabilization of model parameters, burn-in, occurred after 800,000 generations, and every 100th tree after burn-in was used to generate a 50% majority-rule consensus tree. FigTree v1.3.1 [44] was used to construct all trees. Bold branches within trees represent incongruences between the different phylogenetic methods utilized in this study.

Diagnostic PCR

To explore the use of the ITS region as a diagnostic tool for *Sodalis* related bacteria, $ITS^{ala,ile}$ nucleotide alignments were used to identify a *Sodalis* clade specific reverse primer (SgITSR 5'-ACC TTG CAT ATG CCG TCG CT-3'). This oligonucleotide can be used with the 3' end 16S rRNA forward primer (Sg16SF 5'-TGA TTC ATG ACT GGG GTG AA-3') (Figure S1) under the temperature profile of 95 °C for 3 min followed by 35 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, with a final elongation of 72 °C for 5 min. DNA isolated (~300 ng) from various insect hosts were subjected to the diagnostic PCR detection. Negative controls, including *E. coli* and *Bi. tofi*, were included in analyses.

Nucleotide Accession Numbers

The nucleotide sequences from this study have been submitted to the NCBI GenBank database. The 16S rRNA genes (and corresponding accession numbers) used in this study included; *G. brevipalpis* S-symbiont (U64870), *G. pallidipes* S-symbiont (M99060), *G. morsitans* S-symbiont (AY861701), *G. fuscipes* S-symbiont (AY861704), *Sodalis glossinidius* culture (NC_007712), *Cr. melbae* symbiont (EF174495), *Eu. grandis* S-symbiont (AB571330), *Ca. ocellatus* S-symbiont (AB541010), *Sitophilus zeamais* P-symbiont (AF548140, AF548141), *Si. oryzae* P-symbiont (AF548138, AF548139), *G. brevipalpis* P-symbiont (NC_004344), *Cu. sikkimensis* S-symbiont (AB517595), *Bi. tofi* (AM774412), *Yersinia pestis* (NC_003143), *Salmonella enterica* (NC_003198), *E. coli* (NC_000913), *Erwinia amylovora* (NC_013961), *Pantoea vagans* (NC_014562), *Vibrio fischeri* (NC_006840), *Pseudomonas aeruginosa* (NC_002516), *Bacillus cereus* (NC_004722), *Ba. subtilis* (NC_000964), and *Ba. pumilus* (NC_009848). The ITS regions (and corresponding accession numbers in the order of ITS^{glu} and $ITS^{ala,ile}$) used in this study included; *So. glossinidius* culture (NC_007712), *Si. oryzae* P-

symbiont (AF548137), *Y. pestis* (NC_003143), *Sa. enterica* (NC_003198), *E. coli* (NC_000913), *Er. amylovora* (NC_013961), *Pa. vagans* (NC_014562), *G. brevipalpis* P-symbiont (NC_004344), *V. fischeri* (NC_006840), *Si. zeamais* P-symbiont (AF548140, AF548141), *Si. oryzae* P-symbiont (AF548138, AF548139), *Ps. aeruginosa* (NC_002516), *Ba. cereus* (NC_004722), *Ba. subtilis* (NC_000964), and *Ba. pumilus* (NC_009848).

RESULTS AND DISCUSSION

Amplification of ITS Regions

The annotated *Sodalis* genome contains 2 distinct ITS1 regions [45]; a 671 bp ITS which encodes both tRNA-ala and tRNA-ile ($ITS^{ala,ile}$) and an additional 492 bp ITS region containing tRNA-glu (ITS^{glu}). Although multiple copies are found throughout the genome, no sequence divergence is observed within ITS regions due to the pervasiveness of concerted evolution in the rRNA operon [45,46]. In contrast, the genome of the *G. brevipalpis* [47] and *G. morsitans* [48] P-symbiont *Wigglesworthia* retains only two copies of an ITS1 region encoding only tRNA-glu (ITS^{glu}), consisting of 270 bp or 225 bp, respectively, with no intragenomic nucleotide sequence variation and an intergenomic nucleotide sequence identity of 63.7%. The primers used in this study were designed to be specific to *Sodalis* and did not amplify the *Wigglesworthia* ITS region (Figure S2).

Upon sequencing of the ITS regions, ranges in both size (Table 1) and intra- and inter-genomic variation (Table 2) were observed in both $ITS^{ala,ile}$ and ITS^{glu} regions for the examined microbes. Interestingly, the chestnut weevil *Cu. sikkimensis* S-symbiont isolate only amplified one PCR product, with an $ITS^{ala,ile}$ not detected. ITS variation has been linked to functional divergence and differences in ecological capabilities in bacteria [49-51], whether the lack of amplification of the $ITS^{ala,ile}$ from the *Cu. sikkimensis* S-symbiont represents an adaptive

response to particularities of that symbiotic lifestyle remains unclear. Lastly, the free-living *Bi. tofi* amplified two distinct intragenomic ITS^{glu} regions with the highest intragenomic diversity (86.1%–87.5%) observed within this study (Table 2). The amplification of two distinct ITS^{glu} regions by the free-living *Bi. tofi* may represent variation found in the ancestral lineage, which has been purged within the symbionts. In support, *E. coli* also exhibits a similar trend by encoding four ITS^{glu} copies within its genome, which can be divided into two groups, ranging in nucleotide sequence identity from 88.2%–99.2%. It is also tempting to note that *Bi. tofi* was isolated from the biofilm of a tufa deposit [38] which would have increased exposure to the introduction of foreign DNA, potentially contributing to ITS^{glu} variation. Contrastingly, horizontal transfer events are thought to be negligible in the evolution of endosymbionts due to their intracellular localization and reduced recombination rates [10].

ITS Sequence Variation and Molecular Systematics of Sodalis-Allied Symbionts

The ITS sequences, originating from insects harboring *Sodalis* and allied bacteria, were subject to molecular phylogenetic analyses. When examining the ITS^{glu} and $ITS^{ala,ile}$ regions, there was a range of conservation throughout the sequences. Due to functional constraint associated with the tRNA genes, *Sodalis* and related bacterial sequences shared close to 100% sequence identity, with the exception of a low number of point mutations (*i.e.*, <5 between different isolates). Additional conserved motifs, within both ITS regions, were the box A anti-terminator sequence for RNA transcription [52], where all *Sodalis* and related bacteria encoded an identical sequence (5'-CGCTCTTTAACAAT-3') and the RNAse III recognition sites located proximal to the 3' end of the 16S rRNA gene and the 5' end of the 23S rRNA gene [53].

To determine the utility of the 16S rRNA + ITS regions as a tool for resolving relationships and understanding the degree of diversity between *Sodalis* and allied symbionts, NJ, MP and

Bayesian phylogenetic analyses were performed. The resulting phylogenetic trees of 16S rRNA + ITS^{glu} and 16S rRNA + $ITS^{ala,ile}$ (Figures 1 and 2, respectively) gave substantially the same topology and were generally concordant with 16S rRNA based phylogeny [29,32], yet provided stronger resolution among the *Sodalis* and allied bacteria as indicated with relatively higher MP bootstrap (BS) and Bayesian posterior probability (PP) support for most nodes. Phylogenetic analyses of ITS based trees reflect the conserved nature of ITS regions within tsetse isolates (Figures 1 and 2), with both ITS^{glu} and $ITS^{ala,ile}$ trees containing a well-supported monophyletic nest of *Sodalis* isolates, distinct from other *Enterobacteriaceae*, and suggestive of diversification potentially attributed to tsetse host adaptation. Increased sequence divergence of $ITS^{ala,ile}$ with *Sodalis* isolates from *G. pallidipes* and *G. brevipalpis* hosts was also observed, although BS and PP values were not robust at this node.

Within the *Sodalis*-like symbiont clade, the *Sitophilus* P-symbiont $ITS^{ala,ile}$ sequences also displayed significant variation from the remaining insect symbiont sequences, resulting in their own clade with high MP BS and Bayesian PP support (Figure 2). Contrastingly, 16S rRNA based phylogenies intertwine the symbionts from various *Sitophilus* hosts [32], due to rRNA heterogeneities within a genome, most likely arising from a reduction in the efficacy of recombinational gene conversion due to the loss of associated DNA repair loci [54]. Moreover, *Cr. melbae*, *Eu. grandis*, and *Ca. ocellatus* symbionts group together with high support, in both phylogenies despite being housed in insects of two different taxonomic orders, suggesting a recent establishment within each host from a common environmental progenitor and/or possible horizontal transfer of symbionts. The infection of *Sodalis*-like bacteria has been reported from only a minority of populations with low frequency in both *Ca. ocellatus* [24] and *Eu. grandis* [29], this erratic distribution further supports relatively recent host establishments. Displaying

similarities in their infection patterns, the aphid S-symbionts *Candidatus* Hamiltonella defensa and *Candidatus* Regiella insecticola have been shown to establish within phylogenetically diverse hosts [55]. A similar phylogenetic pattern has also been described for the monophyletic *Arsenophonus* genus where some of the symbionts display parallel evolution with their hosts while others demonstrate haphazard association with distant host taxa ranging from insects to plants [30]. Furthermore, the internal node depicting the most recent common ancestor of *Bi. tofi* and the *Sodalis*-allied bacteria, within both the 16S rRNA + ITS^{glu} and 16S rRNA + $ITS^{ala,ile}$ phylogenies, represents inferred lineage splitting that gives rise to symbiotic and free-living sister groups. The transition into symbiosis by the *Sodalis*-allied bacteria appears to have occurred following the diversification of the environmental *Bi. tofi*. Lastly, combining both 16S rRNA and ITS^{glu} regions in our molecular phylogenetic analyses, proved useful towards resolving the taxonomic placement of the *Cu. sikkimensis* S-symbiont. Previously, the phylogenetic placement of this symbiont, based on either the 16S rRNA [24,26,32] or the *groEL* [26] gene, had remained uncertain with low support for grouping with *Sodalis*. Upon utilizing both 16S rRNA and ITS^{glu} regions, the *Cu. sikkimensis* S-symbiont lineage was placed outside of the *Sodalis*-allied symbiont/*Bi. tofi* clade with strong statistical support (Figure 1).

Diagnostic PCR Detection of Sodalis-Like Symbiotic Bacteria

To aid in the detection of *Sodalis*-allied bacteria in novel insect hosts, clade specific ITS primers were synthesized. Using this primer set, with the exception of *Cu. sikkimensis* which appears not to encode an $ITS^{ala,ile}$ region, amplicons were consistently detected in all insect hosts from this study (Figure 3). This primer set was specific to symbiotic *Sodalis*-allied bacteria and did not amplify the free-living relative *Bi. tofi*, *Cu. sikkimensis* S-symbiont, and *E. coli* isolates. We propose the use of this oligonucleotide set as a diagnostic marker for the identification of

additional *Sodalis*-allied symbionts in the field.

Potential Implications for Host Acquisition by Symbionts

Symbiosis is a significant component in the ecology of many microbes and insects in nature. Likewise, the origins of bacterial symbioses are tremendously diverse, ranging from evolutionary transitions between various host associations and environmental lifestyles [56]. Our results support that the infection of *Sodalis*-like bacteria have evolved repeatedly, through multiple opportunities, in a wide array of insect lineages. High nucleotide similarity in the ITS regions among isolates from diverse insect hosts (*i.e*., hippoboscid, shieldbug and stinkbugs) may suggest horizontal transfer among insect species, or establishment by a free-living generalist with an enhanced capability to infect a broad range of insect hosts coupled with insufficient time for diversification. Other symbionts specifically, *Sodalis* and *Sitophilus* symbionts within tsetse and weevil hosts respectively, demonstrate clear separation from other Enterobacteriaceae indicating sufficient association time to allow for diversification of the examined ITS regions. Symbionts of recent origin are believed to be potential sources of novel traits, contrary to P-symbionts which are incapable of such due to genome degradation and secluded host intracellular localization (conferring protection from host immunological defenses but also shielding these microbes from acquiring new genes through horizontal transfer) resulting from extensive host co-evolution.

We speculate that the radiation of *Sodalis*-like bacteria into a diverse range of insects may follow the evolutionary source-sink model [57]. This model illustrates possible events in the early and intermediate stages of establishment into novel habitats, where an evolutionarily stable reservoir (*i.e*., source), has members that migrate from the population into relatively unstable habitats (*i.e*., sinks). Once in a sink, the population faces new challenges, such as host immune

defenses or competition with resident microorganisms. In some cases, continuous emigration from the reservoir may enable adaptive evolution within the population and possibly transform the sink into a new source, able to persist and maintain throughout generations of its host, as a self-sustaining population. The symbiotic association of *Sodalis* with tsetse may be an example of a sink that has evolved into a source, whereby symbiont localization in the milk glands [58,59] (an organ used to feed tsetse larval instars during in utero development), now ensures vertical transmission to future generations of tsetse hosts. The source-sink model of evolution, although traditionally associated with pathogen emergence [60,61], may also prove beneficial towards our discussion on the evolution of symbiosis. Additional studies are needed to demonstrate if positive population growth persists through host reproduction in other insect hosts and to determine the mechanisms enabling symbiont transmission.

The recent discoveries within diverse insects of bacteria closely related to *Sodalis*, raises many experimentally approachable questions, with arguably the most significant being the characterization of conferred benefits and contributory roles towards host phenotypes. The molecular diagnostic markers proposed in this study will facilitate additional identification of related microbes in novel hosts, which will increase the number of symbioses available for comparative genomic and functional studies that aim to elucidate the reciprocal adaptations arising from symbiosis. By integrating into different host backgrounds, the outcomes of the symbioses are likely to not only be varied, but also significantly affect both partners due to tailoring in response to differences in host ecology and physiology.

CONCLUSIONS

This study reports the utility of the ITS region as a tool for both identification and enhanced resolution of the diversity associated with the ever increasing *Sodalis* allied insect

symbiont clade. The similar ITS sequences observed among the tsetse *Sodalis* isolates support previous research describing its lack of divergence between tsetse species [20,32,62]. Importantly, the ITS genomic regions were able to further resolve the relatedness of *Sodalis*-allied bacteria and group the insect host associated bacteria distinct from environmental relatives, providing evidence for its use in future investigations. Utilizing genomic regions, such as surface encoding genes, which may evolve to adapt to specific host backgrounds [32,63], along with ITS regions, with its increased evolutionary rate in comparison to the adjacent 16S rRNA gene [33], as tools to understand the evolution of and ecological adaptations made by symbiotic bacteria will enhance the understanding of steps during symbiont transition. Additionally, the use of *Sodalis*-clade specific primers described in this study provides a diagnostic tool that will aid in the rapid detection of members of this group in field studies within novel insect hosts, further facilitating comparative studies which aim to characterize the reciprocal adaptations involved in different symbioses.

REFERENCES

1. Douglas, A.E. *The Symbiotic Habit*; Princeton University Press: Princeton, NJ, USA, 2010.
2. Buchner, P. *Endosymbiosis of Animals with Plant Microorganisms*; John Wiley and Sons: New York, NY, USA, 1965.
3. Chen, X.; Song, L.; 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*, 49-58.
4. Clark, M.A.; Moran, N.A.; Baumann, P.; Wernegreen, J.J. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* **2000**, *54*, 517-525.
5. Montllor, C.B.; Maxmen, A.; Purcell, A.H. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. *Ecol. Entomol*. **2002**, *27*, 189-195.
6. Oliver, K.M.; Russell, J.A.; Moran, N.A.; Hunter, M.S. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1803-1807.
7. Tsuchida, T.; Koga, R.; Fukatsu, T. Host plant specialization governed by facultative symbiont. *Science* **2004**, *303*, 1989.
8. Jeyaprakash, A.; Hoy, M.A. Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* **2000**, *9*, 393-405.
9. Wernegreen, J.J. Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet*. **2002**, *3*, 850-861.

10. Moran, N.A.; McCutcheon, J.P.; Nakabachi, A. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet*. **2008**, *42*, 165-190.
11. Rio, R.V.M.; Hu, Y.; Aksoy, S. Strategies of the home-team: Symbioses exploited for vector-borne disease control. *Trends Microbiol*. **2004**, *12*, 325-336.
12. Aksoy, S. Tsetse—A haven for microorganisms. *Parasitol. Today* **2000**, *16*, 114-118.
13. Aksoy, S.; Weiss, B.; Attardo, G. Paratransgenesis applied for control of tsetse transmitted sleeping sickness. *Adv. Exp. Med. Biol*. **2008**, *627*, 35-48.
14. Dale, C.; Maudlin, I. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans. Int. J. Syst. Bacteriol.* **1999**, *49*, 267-275.
15. Aksoy, S.; Pourhosseini, A.A.; Chow, A. Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to Enterobacteriaceae. *Insect Mol. Biol.* **1995**, *4*, 15-22.
16. Cheng, Q.; Ruel, T.D.; Zhou, W.; Moloo, S.K.; Majiwa, P.; O'Neill, S.L.; Aksoy, S. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. *Med. Vet. Entomol*. **2000**, *14*, 44-50.
17. Welburn, S.C.; Maudlin, I.; Ellis, D.S. *In vitro* cultivation of rickettsia-like organisms from *Glossina* spp. *Ann. Trop. Med. Parasitol*. **1987**, *81*, 331-335.
18. Farikou, O.; Njiokou, F.; Mbida Mbida, J.A.; Njitchouang, G.R.; Djeunga, H.N.; Asonganyi, T.; Simarro, P.P.; Cuny, G.; Geiger, A. Tripartite interactions between tsetse flies, *Sodalis glossinidius* and trypanosomes—An epidemiological approach in two historical human African trypanosomiasis foci in Cameroon. *Infect. Genet. Evol*. **2010**, *10*, 115-121.

19. Lindh, J.M.; Lehane, M.J. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. *Antonie Van Leeuwenhoek* **2011**, *99*, 711-720.

20. Aksoy, S.; Chen, X.; Hypsa, V. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). *Insect Mol. Biol*. **1997**, *6*, 183-190.

21. Toh, H.; Weiss, B.L.; Perkin, S.A.; Yamashita, A.; Oshima, K.; Hattori, M.; Aksoy, S. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res*. **2006**, *16*, 149-156.

22. Snyder, A.K.; Deberry, J.W.; Runyen-Janecky, L.; Rio, R.V.M. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. Biol. Sci.* **2010**, *277*, 2389-2397.

23. Warnecke, F.; Luginbühl, P.; Ivanova, N.; Ghassemian, M.; Richardson, T.H.; Stege, J.T.; Cayouette, M.; McHardy, A.C.; Djordjevic, G.; Aboushadi, N.; *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **2007**, *450*, 560-565.

24. Kaiwa, N.; Hosokawa, T.; Kikuchi, Y.; Nikoh, N.; Meng, X.Y.; Kimura, N.; Ito, M.; Fukatsu, T. Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. *Appl. Environ. Microbiol*. **2010**, *76*, 3486-3494.

25. Grünwald, S.; Pilhofer, M.; Höll, W. Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles (*Coleoptera: Cerambycidae*). *Syst. Appl. Microbiol.* **2010**, *33*, 25-34.

26. Toju, H.; Hosokawa, T.; Koga, R.; Nikoh, N.; Meng, X.Y.; Kimura, N.; Fukatsu, T. "*Candidatus* Curculioniphilus buchneri", a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. *Appl. Environ. Microbiol*. **2010**, *76*, 275-282.

27. Fukatsu, T.; Koga, R.; Smith, W.A.; Tanaka, K.; Nikoh, N.; Sasaki-Fukatsu, K.; Yoshizawa, K.; Dale, C.; Clayton, D.H. Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. *Appl. Environ. Microbiol*. **2007**, *73*, 6660-6668.

28. Nováková, E.; Hypsa, V. 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*, 131-135.

29. Kaiwa, N.; Hosokawa, T.; Kikuchi, Y.; Nikoh, N.; Meng, X.Y.; Kimura, N.; Ito, M.; Fukatsu, T. Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera: Scutelleridae). *Zool. Sci.* **2011**, *28*, 169-174.

30. Novakova, E.; Hypsa, V.; Moran, N.A. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol.* **2009**, *9*, 143.

31. Naum, M.; Brown, E.W.; Mason-Gamer, R.J. Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the Enterobacteriaceae? *J. Mol. Evol.* **2008**, *66*, 630-642.

32. 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*. **2011**, *317*, 143-151.

33. Gürtler, V.; Stanisich, V.A. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **1996**, *142*, 3-16.

34. Brown, M.V.; Fuhrman, J.A. Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat. Microb. Ecol*. **2005**, *41*, 15-23.

35. Erwin, P.M.; Thacker, R.W. Cryptic diversity of the symbiotic cyanobacterium *Synechococcus spongiarum* among sponge hosts. *Mol. Ecol*. **2008**, *17*, 2937-2947.

36. Holmes, D.S.; Bonner, J. Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* **1973**, *12*, 2330-2338.

37. Aoki, K.; Kato, M.; Murakami, N. Glacial bottleneck and postglacial recolonization of a seed parasitic weevil, *Curculio hilgendorfi*, inferred from mitochondrial DNA variation. *Mol. Ecol*. **2008**, *17*, 3276-3289.

38. Verbarg, S.; Frühling, A.; Cousin, S.; Brambilla, E.; Gronow, S.; Lünsdorf, H.; Stackebrandt, E. *Biostraticola tofi* gen. nov., spec. nov., a novel member of the family *Enterobacteriaceae. Curr. Microbiol*. **2008**, *56*, 603-608.

39. Rozen, S.; Skaletsky, H.J. Primer3 on the WWW for General Users and for Biologist Programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology*; Krawetz, S., Misener, S., Eds.; Humana Press: Totowa, NJ, USA, 2000; pp. 365-386.

40. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **2004**, *5*, 113.

41. Swofford, D.L. *PAUP 4.0-Phylogenetic Analysis Using Parsimony*, Version 4; Sinauer Associates: Sunderland, MA, USA, 2002.

42. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572-1574.

43. Nylander, J.A.A. *MrModeltest v2*; Evolutionary Biology Centre, Uppsala University: Uppsala, Sweden, 2004.

44. FigTree. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 15 November 2011).
45. Stewart, F.J.; Cavanaugh, C.M. Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria. *J. Mol. Evol*. **2007**, *65*, 44-67.
46. Liao, D. Gene conversion drives within genic sequences: Concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol*. **2000**, *51*, 305-317.
47. Akman, L.; Yamashita, A.; Watanabe, H.; Oshima, K.; Shiba, T.; Hattori, M.; Aksoy, S. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia. Nat. Genet*. **2002**, *32*, 402-407.
48. Rio, R.V.M.; Symula, R.E.; Wang, J.; Lohs, C.; Wu, Y.; Snyder, A.K.; Bjornson, R.D.; Oshima, K.; Biehl, B.S.; Perna, N.T.; *et al*. Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *mBio* **2012**, *3,* e00240-11.
49. Rocap, G.; Distel, D.L.; Waterbury, J.B.; Chisholm, S.W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol*. **2002**, *68*, 1180-1191.
50. Jaspers, E.; Overmann, J. Ecological significance of microdiversity: Identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl. Environ. Microbiol*. **2004**, *70*, 4831-4839.
51. Hahn, M.W.; Pöckl, M. Ecotypes of planktonic actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. *Appl. Environ. Microbiol*. **2005**, *71*, 766-773.

52. Berg, K.L.; Squires, C.; Squires, C.L. Ribosomal RNA operon anti-termination. Function of leader and spacer region box B-box A sequences and their conservation in diverse micro-organisms. *J. Mol. Biol.* **1989**, *209*, 345-358.

53. Srivastava, A.K.; Schlessinger, D. Mechanism and regulation of bacterial ribosomal RNA processing. *Annu. Rev. Microbiol*. **1990**, *44*, 105-129.

54. Dale, C.; Wang, B.; Moran, N.; Ochman, H. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Mol. Biol. Evol*. **2003**, *20*, 1188-1194.

55. Russell, J.A.; Latorre, A.; Sabater-Munoz, B.; Moya, A.; Moran, N.A. Side-stepping secondary symbionts: Widespread horizontal transfer across and beyond the Aphidoidea. *Mol. Ecol*. **2003**, *12*, 1061-1075.

56. Sachs, J.L.; Skophammer, R.G.; Regus, J.U. Evolutionary transitions in bacterial symbiosis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10800-10807.

57. Pulliam, H.R. Sources, sinks, and population regulation. *Am. Nat.* **1988**, *132*, 652-661.

58. Ma, W.C.; Denlinger, D.L. Secretory discharge and microflora of milk gland in tsetse flies. *Nature* **1974**, *247*, 301-303.

59. Attardo, G.M.; Lohs, C.; Heddi, A.; Alam, U.H.; Yildirim, S.; Aksoy, S. Analysis of milk gland structure and function in *Glossina morsitans*: Milk protein production, symbiont populations and fecundity. *J. Insect Physiol*. **2008**, *54*, 1236-1242.

60. Fisher, M.C. Endemic and introduced haplotypes of *Batrachochytrium dendrobatidis* in Japanese amphibians: Sink or source? *Mol. Ecol*. **2009**, *18*, 4731-4733.

61. Rambaut, A.; Pybus, O.G.; Nelson, M.I.; Viboud, C.; Taubenberger, J.K.; Holmes, E.C. The genomic and epidemiological dynamics of human influenza A virus. *Nature* **2008**, *453*, 615-619.

62. Weiss, B.L.; Mouchotte, R.; Rio, R.V.M.; Wu, Y.; Wu, Z.; Heddi, A.; Aksoy, S. Interspecific transfer of bacterial endosymbionts between tsetse fly species: Infection establishment and effect on host fitness. *Appl. Environ. Microbiol*. **2006**, *72*, 7013-7021.

63. Weiss, B.L.; Wu, Y.; Schwank, J.J.; Tolwinski, N.S.; Aksoy, S. An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15088-15093.

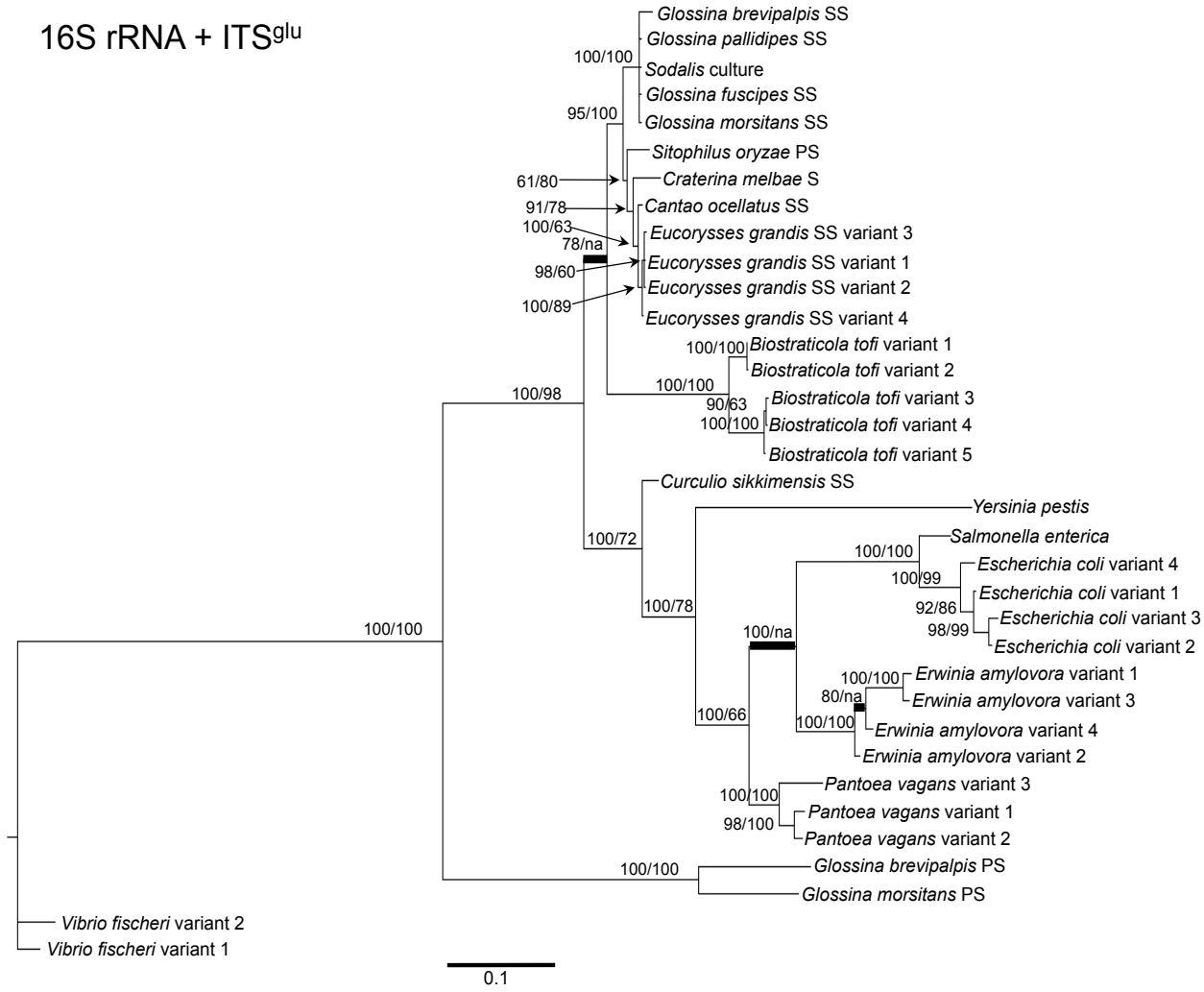


Figure 4.2-1. Phylogenetic placement of *Sodalis* and related symbiotic bacteria within Gammaproteobacteria based on 16S rRNA and ITS^{glu} concatenation. A Bayesian tree, inferred from a total of 2,467 unambiguously aligned nucleotide sites, with support values indicating Bayesian posterior probabilities (PP)/MP bootstrap (BS) is shown. PP indicated as percentage, *i.e.*, PP = 0.95 is depicted as 95. Branches constrained with MP are shown in bold. For insect symbionts, host species are indicated. S = symbiont, SS = S-symbiont, PS = P-symbiont. Scale bar represents substitutions/site.

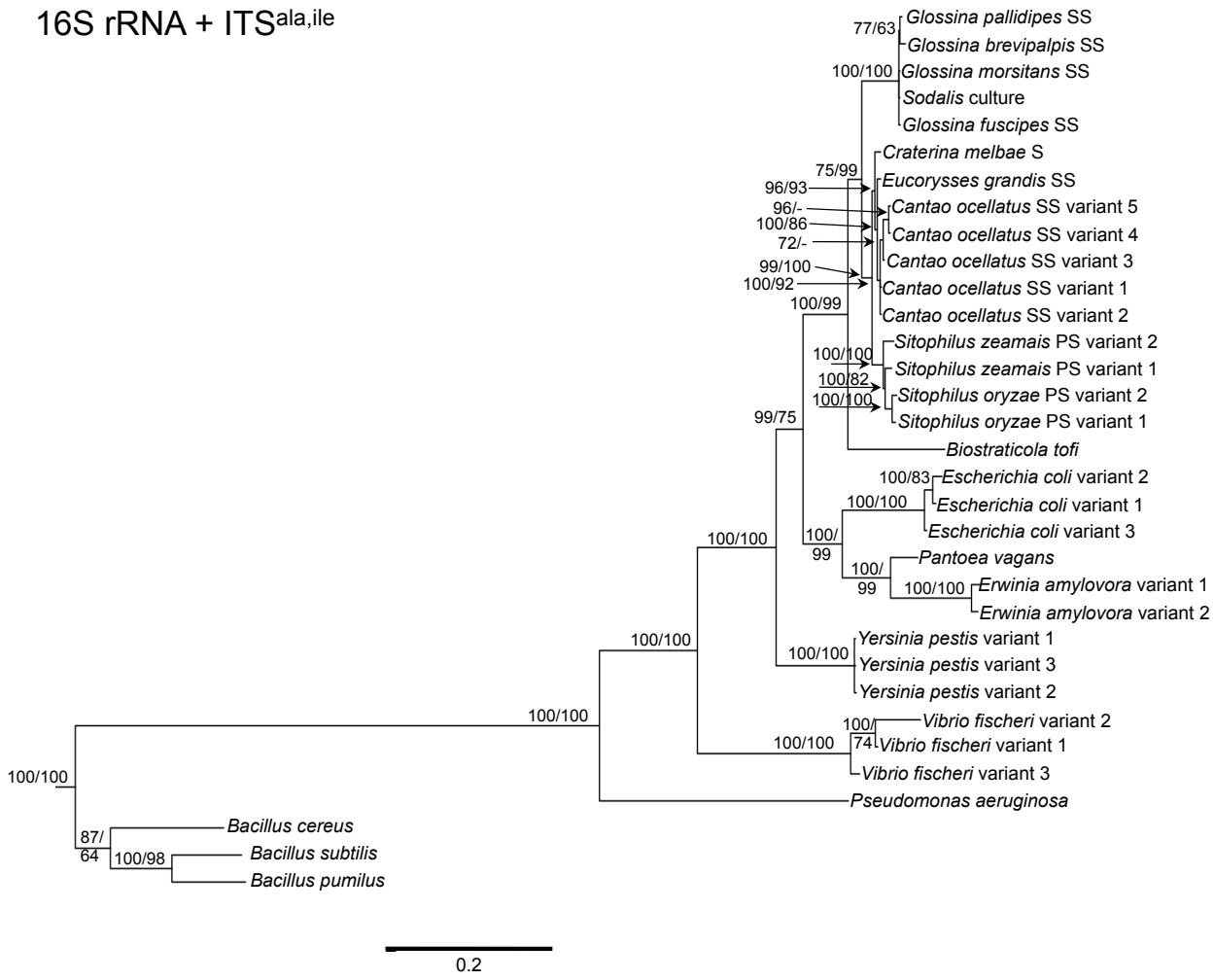


Figure 4.2-2. Phylogenetic placement of *Sodalis* and related symbiotic bacteria within Bacteria based on ITS^{ala,ile}. A Bayesian tree, inferred from a total of 2,762 unambiguously aligned nucleotide sites, with support values indicating Bayesian posterior probabilities (PP)/ MP bootstrap (BS) is shown. PP indicated as percentage, *i.e.*, PP = 0.95 is depicted as 95. Branches collapsed with Bayesian analysis are shown in bold. NJ analysis resulted in a similar phylogeny. For insect symbionts, host species are indicated. S = symbiont,
SS = S-symbiont, PS = P-symbiont. Scale bar represents substitutions/site.

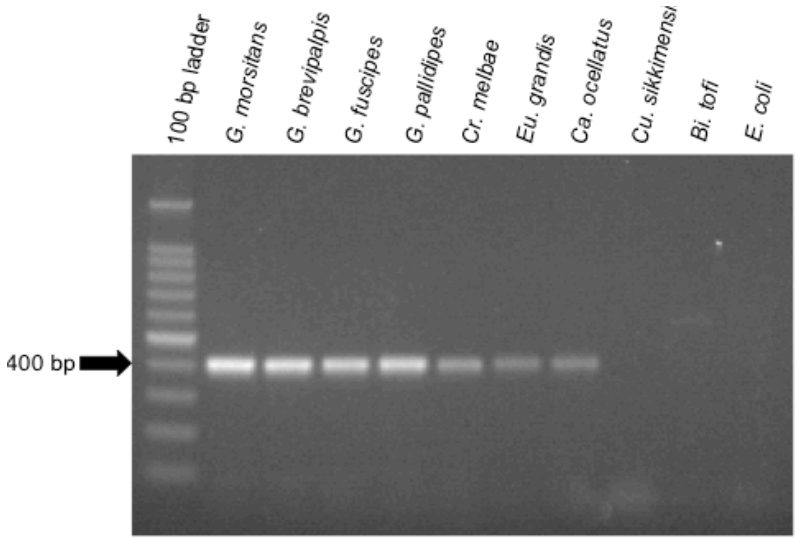


Figure 4.2-3. Diagnostic PCR detection of *Sodalis* and allied insect symbionts using ITS^{ala,ile} specific oligonucleotides and 300 ng of DNA template. An approximately 400 bp product was amplified. Lanes are labeled by either insect host or culture isolate (*i.e.*, *Bi. tofi* and *E. coli*).

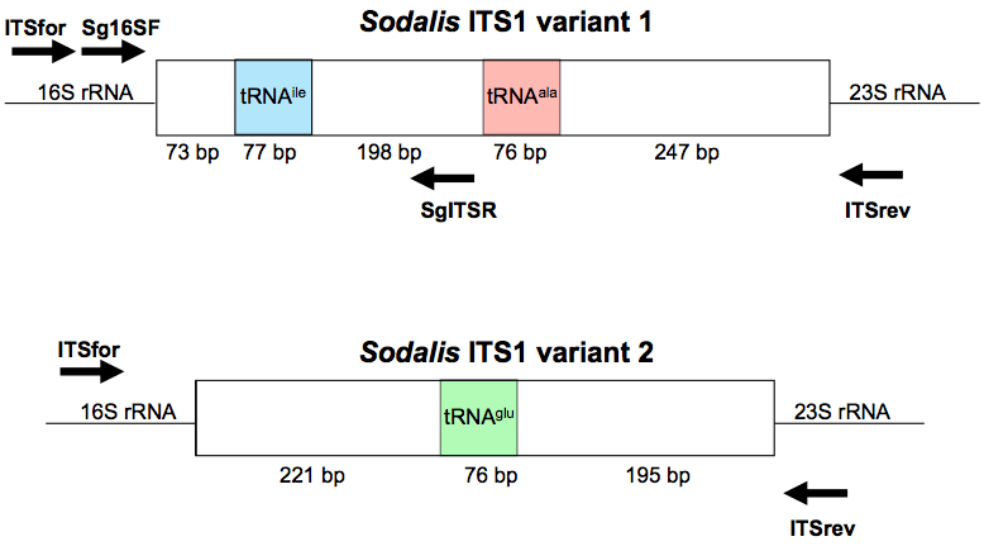


Figure 4.2-S1. Organization of the two *Sodalis* ITS1 variants. Locations of the *Sodalis*-allied symbiont specific PCR primers (Sg16SF and SgITSR) and ITS sequencing primers (ITSfor and ITSrev) are indicated.

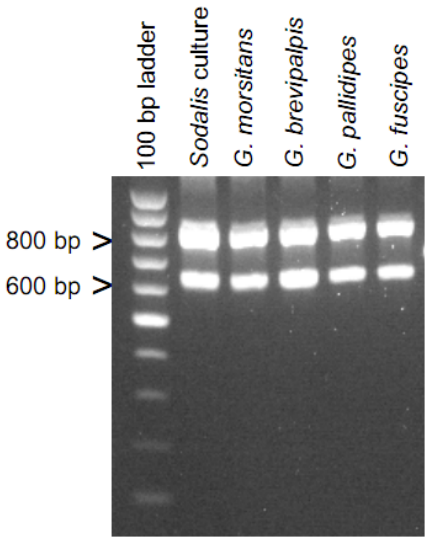


Figure 4.2-S2. Specificity of ITS primers for *Sodalis*. The ITSfor and ITSrev primers do not amplify the corresponding *Wigglesworthia* ITS^{glu}. Both *Sodalis* ITS1 variants were amplified; the 646 bp amplicon contains the 492 bp ITS^{glu}, while the 825 bp amplicon contains the 671 bp $ITS^{ala,ile}$. Tsetse host species are indicated.

Table 4.2-1. Comparison of the ITS^{glu} and ITS^{ala,ile} lengths for *Sodalis* and allied symbionts. With the exception of *Sodalis* from culture, host species are indicated.

	ITS^{glu}	**ITS^{ala,ile}**
	Length (bp)	**Length (bp)**
G. brevipalpis SS	492	671
G. fuscipes SS	492	671
G. pallidipes SS	492	671
G. morsitans SS	492	671
Sodalis culture	492	671
Si. oryzae PS	668	836, 837
Si. zeamais PS	n/a	837
Cr. melbae S	544	861
Ca. ocellatus SS	693	861
Eu. grandis SS	463	861
Cu. sikkimensis SS	307	Not detected
Bi. tofi (tufa deposit)	604, 614	796

S: symbiont; PS: primary symbiont; SS: secondary symbiont.

Table 4.2-2. Percent nucleotide identity of ITS regions among *Sodalis* and allied symbionts. Host genera are specified, with the exception of the free-living *Biostraticola* genus.

ITS^{glu} % identity							
Glossina	*Sitophilus*	*Curculio*	*Craterina*	*Cantao*	*Eucorysses*	*Biostraticola*	
98.6–100	94.6–96.0	69.2–69.8	93.3–93.9	94.6–96.0	94.5–96.7	76.6–79.2	*Glossina*
	100	68.9	90.8	93.7	96.8–97.5	73.9–75.1	*Sitophilus*
		100	68.8	70.3	70.5–71.2	67.1–68.0	*Curculio*
			100	95.3	93.6–94.2	74.2–79.2	*Craterina*
				100	98.7–99.4	73.0–76.8	*Cantao*
					99.4–100	79.8–81.3	*Eucorysses*
						86.1–100	*Biostraticola*

ITS^{ala,ile} % identity						
Glossina	99.4–100					
Sitophilus	86.3–87.3	97.5–100				
Craterina	87.0–87.7	94.9–95.7	100			
Cantao	84.6–87.2	94.4–97.8	98.0–99.1	99.3–100		
Eucorysses	86.8–87.3	94.7–95.4	98.5	98.7–99.3	100	
Biostraticola	74.6–75.1	73.6–73.9	75.0	68.4–74.9	74.4	100
	Glossina	*Sitophilus*	*Craterina*	*Cantao*	*Eucorysses*	*Biostraticola*

Chapter 5: Concluding remarks

Most if not all, organisms have formed symbioses with more than one other species. The significance of these relationships is continually being described, with microbial symbionts demonstrated to play essential roles in all facets of host biology, ranging from ecology to human health (Reviewed in Hussa and Goodrich-Blair 2013). The majority of all animals and plants contain highly complex microbial communities, including humans. To understand how a highly diverse microbiota is formed and persists harmoniously within a host, I examined the genomic evolution and interactions of a more simple multipartite system, composed of the tsetse fly host (*Glossina morsitans morsitans*), and its microbial partners; *Wigglesworthia*, *Sodalis*, *Wolbachia* and *Trypanosoma brucei* subspp. This work expands our understanding of the evolution towards symbiont specialization and provides insights into how partner interactions promote the formation and stability of a holobiont (i.e. the macroscopic host and associated microbes).

First, I examined how metabolic interplay is important for maintaining a stable symbiotic environment, providing a means for a recently established microbial symbiont to adapt into co-residence. These studies described how *Wigglesworthia* likely provisions thiamine monophosphate to the tsetse host, in a regulated manner, which may be obtained by *Sodalis* (who lacks this biosynthetic capability) through a functionally intact thiamine transporter. This nutrient is required by *Sodalis* for growth and influences its within host population size. The stable environment and successful co-residence within tsetse is therefore created from the physiological balance of the species involved, which includes both transcriptional and symbiont density regulation. This work demonstrates that microbial genomes begin to evolve early upon establishment within a symbiotic community to form interdependent collaborations, accommodating their own requirements as well as the needs of the symbiotic system. Future

work on this interaction should examine whether the changes in *Wigglesworthia* population density and biosynthetic loci transcription are host or symbiont mediated. Additionally, studies should compare genomic modifications of multiple *Sodalis* isolates and examine its evolution *in vivo* to determine whether the forces driving the formation of metabolic interactions within this system are a result of an excess of specific available nutrients (as described in the Black Queen Hypothesis (Morris *et al.* 2012)) and/or the random loss of symbiont capabilities caused by genetic drift. Understanding how new members forge relationships within a previously established system help describe initial steps taken in the evolution of a successful holobiont. Metabolic integration of symbionts, acquired via genome complementation, may prove pivotal towards enhancing efficiency and stability of the microbiota.

The strict symbiotic lifestyle is associated with the retention of only the genomic information that is vital for the maintenance of the mutualism. My research was the first to demonstrate how small genetic differences among closely related ancient symbionts has functional significance, accommodating the physiological needs and ecology of their host. This work described how the chorismate and folate biosynthetic capabilities, encoded by Wgm and not Wgb, remain functional and maintain an active role in the biology of *G. morsitans*, particularly during blood meal digestion, progeny development and trypanosome challenge. Future studies should begin to compare other closely related symbionts, to determine whether differences in host phenotypic traits, such as diet choice or environmental preference (temperature, salinity, water availability, etc.), are due to distinctive microbial capabilities. These studies may elucidate novel modes of pest control, by targeting symbiont contributions, and identify new ways to enhance the fitness of beneficial organisms facing environmental challenges.

With specific regards to tsetse, distinctions among *Wigglesworthia* spp. capabilities may be a contributing factor to the differences observed in tsetse vector competence (Harley 1971, Moloo and Kutuza 1988, Moloo *et al.* 1994). Future studies should further examine the role of Wgm chorismate and folate biosynthetic capabilities during trypanosome infection by inhibiting the pathway and subsequently observing the progression of trypanosomes within *G. morsitans*. Additionally, *G. brevipalpis*, a tsetse species with relatively lower vector competence (Harley 1971, Moloo and Kutuza 1988, Moloo *et al.* 1994), may be given folic acid supplemented blood meals to determine whether this excess nutrient could increase the frequency of mature trypanosome infections. Lastly, investigating the potential link between the heightened immune response and specific immune factors (e.g. ROS) with the production of folate by Wgm will provide deeper insight into *G. morsitans* biology. Enhanced knowledge of specific *Wigglesworthia* contributions may be capitalized to create novel and directed control strategies aimed at tsetse host species that are the major disease vectors.

While comparison of *Wigglesworthia* spp. has enabled the discovery of unique capabilities among ancient mutualists, the annotated genome of *Sodalis* facilitated investigations regarding initial genomic alterations occurring in light of novel host establishments. By sequencing genomic regions traditionally associated with an accelerated evolutionary rate (i.e. internal transcribed spacer (ITS) regions and genes encoding outer membrane proteins) from *Sodalis* and closely related bacteria, the phylogenetic resolution of this bacterial clade was enhanced and supports the hypothesis that diversification occurs due to host adaptation. Additionally, this research demonstrated that some outer membrane proteins evolve within the context of novel hosts and may be important for establishing initial relationships, while others do not appear to be involved in adapting to new environments, possibly due to conserved roles.

These studies, combined with a recent report of strain HS (Clayton *et al.* 2012), suggest that this bacterial clade has an enhanced ability to form multiple, independent symbiotic associations and originates from an environmental progenitor. Identifying the qualities within this bacterial group that have enabled many disparate associations may divulge molecular mechanisms providing these opportunities and will contribute to our basic understanding of microbial symbiont establishment. Future studies could further examine the surface encoding loci shown to exhibit genetic divergence, to determine whether they provide a fitness advantage within the novel host, such as aid in evading the host immune system. Genes encoding outer membrane proteins of interest could be transformed into and produced by *Sodalis* or related bacteria to examine whether these components are sufficient to enable successful establishment within a specific host. Examining the genetic adaptations that occur during integration of microbes with macroscopic hosts provides insight towards fundamental evolutionary aspects involved in symbiosis.

The overall findings of this research advance our knowledge of genomic adaptations that occur during different stages of symbiotic integration and the evolution of a holobiont. First, a deeper understanding of the process of partner inclusion within a symbiosis was obtained. Mechanisms developed to minimize competition, within the tsetse host, include nutritional interdependence and complementary loss of biosynthetic capabilities by the recently established member. Additionally, I demonstrate that the evolution of ancient bacterial genomes may result in the retention of unique capabilities (despite a high degree of genomic synteny) that can have a significant impact in the success of a holobiont, contributing to host physiology and phenotypic variation. Lastly, alterations in symbiont loci occur upon novel insect host establishment, possibly enabling the formation of initial relationships.

Microbial symbionts are integrated into many basic biological functions of their macroscopic hosts, such as the maintenance of health through nutrient provisioning and immune stimulation (Reviewed in Hussa and Goodrich-Blair 2013). Understanding how these relationships develop throughout time is pivotal to obtaining a more holistic view of the evolution of biological systems. Taken together, these studies contribute to our knowledge of host-associated microbial community dynamics.

REFERENCES

1. Hussa EA & Goodrich-Blair H (2013) It takes a village: Ecological and fitness impacts of multipartite mutualism. Annu Rev Microbiol 67:161-178.

2. Morris JJ, Lenski RE, & Zinser ER (2012) The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio 3(2):e00036-00012.

3. Harley JM (1971) Comparison of the susceptibility of infection with *Trypanosoma rhodesiense* of *Glossina pallidipes*, *G. morsitans*, *G. fuscipes* and *G. brevipalpis*. Ann Trop Med Parasitol 65(2):185-189.

4. Moloo SK & Kutuza SB (1988) Comparative study on the susceptibility of different *Glossina* species to *Trypanosoma brucei brucei* infection. Trop Med Parasitol 39(3):211-213.

5. Moloo SK, Kabata JM, & Sabwa CL (1994) A study on the maturation of procyclic *Trypanosoma brucei brucei* in *Glossina morsitans centralis* and *G. brevipalpis*. Med Vet Entomol 8(4):369-374.

6. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C (2012) A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. PLoS Genet 8:e1002990.