Variation in the Structure and Function of Invertebrate-Associated Bacterial Communities

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Frances Blow

September 2016

UNIVERSITY OF LIVERPOOL

Abstract

Microorganisms are intricately involved in the ecology of many insects, often contributing to host fitness and forming evolutionarily stable associations. The interactions between hosts and microbes can significantly alter their evolutionary trajectories, enabling them to adapt to novel environmental conditions. In this thesis I have examined how host ecology can shape the interactions of bacteria with insects of agricultural and epidemiological importance. I have described the bacterial communities associated with *Bactrocera oleae* (the olive fruit fly) and generated draft genome sequences for several members of the gut microbiota, including the symbiotic bacterium "*Candidatus Erwinia dacicola*". Comparative genomic analyses indicate that *Ca.* E. dacicola and a novel facultative bacterium *Tatumella* TA1 may perform key nutritional functions for the host, including the synthesis of essential amino acids and ammonia assimilation from host nitrogenous waste products. *Tatumella* TA1 is consistently associated with all life stages of populations collected in Israel and Crete at low relative abundance, and encodes large adhesion proteins that may assist in attachment to the host epithelium or other members of the microbiota in the *B. oleae* gut. I have also examined the variation in frequency and relative abundance of facultative microbes that infect several *Glossina* spp. (the tsetse fly): the sole vector of African trypanosomes in Sub-Saharan Africa. In addition to three vertically transmitted endosymbionts (*Wigglesworthia*, *Sodalis*, and *Wolbachia*), tsetse flies are infected with two additional potential reproductive manipulators: *Spiroplasma* and *Rickettsia*, and a novel strain of *Klebsiella*. The draft genomes generated for these taxa over the course of this thesis provide the opportunity for future studies in to their role in host biology and how community interactions can shape the transmission and evolutionary dynamics of host-associated microbes.

Acknowledgements

There are a lot of people without whom this thesis would not have been possible, and I am eternally grateful to every single one of them.

First and foremost, I would like to thank my supervisors Alistair and Martha for their help and support over the past four years. Particular thanks go to Alistair, who has given me endless encouragement and mentorship throughout my PhD.

I would like to thank the following people, to whom I am hugely grateful: Ian, who has always been willing to teach and encourage me, and without whom a lot of this work would not have been possible. Arjen, whose encyclopaedic knowledge of the natural world never fails to motivate and inspire me. Greg Hurst for his support and encouragement, and for sharing his extensive knowledge of insects and microbes. Everyone in the Darby lab, and particularly Dave, who has been a constant companion over the course of my PhD. The staff of the CGR, both lab and bioinformatics, who have been incredibly kind and patient, and who are always willing to help. Everyone at Oxitec and the Vontas lab at the University of Crete for their help and advice rearing insects. Natassa, Evgenia, and John Vontas, for making me feel so welcome during my wonderful field season in Crete.

On a personal note, I would like to thank Jess and Kim for giving me the confidence to do a PhD in the first place. Sophie, Chris, Kirsten and Lydia for always being there to support me and make me laugh, along with a list of people that is too long to write here. A special thank you to Mum, Dad, Greg, and Rachel for always having faith in me, and for encouraging me in everything. And finally, thank you to Mike for his unwavering support and generally being wonderful.

Contents

List of Abbreviations

AMP	Antimicrobial peptide
C:N	Carbon to Nitrogen ratio
CI	Cytoplasmic incompatibility
FISH	Fluorescent *in situ* hybridization
fsRIDL	Female-specific release of a dominant lethal
HGT	Horizontal gene transfer
MLST	Multi locus sequence typing
RLO	*Rickettsia*-like organism
ROS	Reactive-oxygen species
SIT	Sterile insect technique
TAGC	Taxon annotated Guanidine Cytosine coverage
WFT	Western Flower Thrip
WHO	World Health Organisation

List of Figures

List of Tables

Chapter 1 : General Introduction

1.1 Symbiotic Bacteria of Insects

Associations with microorganisms are ubiquitous in eukaryotes (Douglas, 2010; McFall-Ngai et al., 2013). Insects form associations with microorganisms that induce diverse phenotypic effects ranging from pathogenicity to obligate mutualism. Such associations have profound effects on the evolutionary trajectories of both parties. By marrying their functional capabilities with those of bacteria, which have fast generation times, opportunities for horizontal transmission, access to additional mobile elements such as plasmids and phage, and adaptations to alternative environments, eukaryotes increase their ability to adapt to novel conditions. These include both abiotic and biotic stressors, and have been evidenced across the great diversity of the *Insecta*. Symbiotic bacteria are able to influence nutrition (Douglas, 2009), immunity (Gross et al., 2009), interactions with natural enemies (Oliver et al., 2003), development (Braendle et al., 2003), and reproduction (Hurst, 2000), and have greatly diversified the phenotypes that we see amongst insects today (Janson et al., 2008).

1.1.1 Symbiont categories

Several classes of bacterial symbiont associate with insects, all of which have broad phylogenetic origins. Nutritional symbionts such as *Buchnera* in aphids or *Tremblaya* in mealybugs are intracellular, vertically transmitted, obligate mutualists that have been co-evolving with their hosts for long periods of time. Both host and bacterium are dependent upon each other for survival in such partnerships, and these relationships are characterised by specialised host organs known as bacteriomes or mycetomes, which accommodate mutualistic bacteria (Baumann, 2005). In contrast to obligate mutualists, where the complementarity of metabolic functions often drives the association, parasitic microbes can form long-term associations with their hosts without contributing to fitness. These are known as reproductive manipulators, the most prominent examples of which are *Arsenophonus*, *Cardinium*, *Rickettsia*, *Spiroplasma* and *Wolbachia*. Reproductive manipulators exploit vertical transmission mechanisms from mother to offspring by biasing sex ratios towards females,

enhancing their own transovarial transmission through female progeny (Hurst, 2000). The persistent nature of these interactions often leads to the attenuation or domestication of the reproductive parasite, as vertical transmission serves to marry together the fate of host and microbe (Chrostek and Teixeira, 2015). Therefore, many are also known to serve mutualistic (Engel and Moran, 2008) functions, particularly in females (Brownlie et al., 2009). The final class of bacterial symbionts associated with insects is the most diverse class: the facultative symbionts. Facultative symbionts are not confined to specialised organs or reproductive tissues so can be both intra and extracellular; they are conditionally beneficial and perform a broader range of functions than obligate primary symbionts; and they are not reliant upon vertical transmission so are free to utilise horizontal transmission or free-living strategies that are independent of the host (Moran et al., 2008). Interactions form a continuum along this range of phenotypes, and due to the highly plastic nature of bacterial genomes and conflict between partners there is considerable potential for fluctuation. In addition, phenotypic effects are heavily reliant upon genotype-genotype interactions between host and bacterium. We are only just beginning to understand the complexity and diversity of interactions that can take place between insects and microbes.

1.1.2 Symbiont habitats

Obligate endosymbionts are often housed in specialised host cells known as bacteriocytes, which facilitate symbiont infection and vertical transmission (Koga et al., 2012). In some insects these cells form an organ called the bacteriome, which is filled with endosymbionts and is tightly regulated by the host to facilitate the establishment and maintenance of symbiosis (Shigenobu and Stern, 2012). In addition to specialised cells, intracellular symbionts can inhabit other host cell types, including those in the reproductive organs, gut, haemolymph and fat body (Pontes and Dale, 2006). Some endosymbionts are highly specialised, such as some strains of *Wolbachia* are only able to inhabit the reproductive organs, whereas others can use multiple cell types throughout the insect body. *Sodalis*, which infects insects from a broad phylogenetic range, is found in the salivary glands, midgut, muscle, reproductive organs, haemolymph, and milk glands of tsetse flies (Cheng and

Aksoy, 1999). Bacteria are also able to colonize extracellular locations within the insect body. The most common of these is the gut, which experiences high exposure to environmental microbes and is a nutrient-rich environment. Gut communities can be transient or highly specialised, and are often shaped by host factors such as diet, gut morphology, and physiological conditions (Engel and Moran, 2008). The insect cuticle can also be colonised by symbiotic bacteria. For example, *Pseudonocardia* bacteria are reared on the cuticle of leaf cutter ants, where they defend their host from parasitic fungi (Currie et al., 2003). Localization has a huge impact upon the evolutionary trajectory of a symbiont, as it strongly influences the physiological environment, transmission mechanisms, and population dynamics experienced.

1.1.3 Adaptation to symbiosis

The lifestyle of a symbiont is reflected in its genomic architecture. Vertically transmitted endosymbionts are exposed to two key evolutionary determinants: genetic drift and the retention of traits that are essential for life within a host (McCutcheon and Moran, 2012). Endosymbiont genomes comprise a subset of genes derived from their free-living ancestors, which have been selected and shaped by life within another organism (Andersson and Kurland, 1998; Morris et al., 2012). They display a distinct set of modifications in comparison to free-living bacteria, including altered codon usage (evident as reduced GC content), rapid protein evolution, and extreme reduction in genome size and content (Moran et al., 2009). Small population size and the lack of opportunity for recombination due to low genetic diversity subjects endosymbionts to Muller's Ratchet, where deleterious mutations accumulate to the point of fixation within a population (Moran, 1996). This leads to thermal instability and reduced efficiency of protein products (Lambert and Moran, 1998; Fares et al., 2002). Gene loss is largely attributed to purifying selection by redundancy with the host cell, and a natural bias towards deletion in prokaryotes (Nilsson et al., 2005). These traits, along with the conflicting fitness priorities of host and symbiont, mean that obligate symbioses can evolve beyond their fitness optima, a phenomenon known as the "evolutionary rabbit hole" (Bennett and Moran, 2015). Whilst these traits may not be particularly beneficial to the

partners of the symbiosis, they have proved indispensable to progress within the field of symbiosis research, particularly in recent years as next generation sequencing technologies have become more accessible. The structure and composition of a symbiont genome allows us to diagnose not just the nature of an association, but also its age, evolutionary trajectory, history, mode of transmission, and the likely number of partners involved.

1.1.4 Alternatives to obligacy

One way of escaping the evolutionary rabbit hole is to adopt an alternative symbiont. There is evidence that this has occurred in several insect clades including weevils (Lefevre et al., 2004), spittlebugs
(Koga et al., 2013; Koga and Moran, 2014), and aphids (Fukatsu and Ishikawa, 1992). Replacing entrenched symbionts represents a risk to the host, as coadaptation creates dependence (Baumann et al., 1997), which could potentially cause disruption if not adequately compensated. However, large fitness benefits stand to be gained from adopting a new symbiont. New symbionts allow hosts to potentially obtain novel functions, and to replace the machinery supplied by existing symbionts with gene products that have not been subjected to Muller's Ratchet for extended periods of time (Koga and Moran, 2014). Mealybugs adopt a similar strategy. Like several other insects, mealybugs form stable associations with at least two partners (Dohlen et al., 2001). The primary symbiont, *Tremblaya princeps*, is housed in the bacteriome (Kantheti et al., 1996), has an incredibly reduced genome that is just 139 kb in size (McCutcheon and Dohlen, 2011), and is the result of a single infection in the ancestor of mealybugs (Husnik and McCutcheon, 2016). This is an ancient association that is supplemented by horizontal gene transfers of bacterially derived genes, from taxa other than *Tremblaya*, in to the host genome, which are over expressed in the bacteriocytes and compensate for the extreme reduction of *Tremblaya* (Husnik et al., 2013; Husnik and McCutcheon, 2016). Within the *Tremblaya* cytoplasm is housed another symbiotic bacterium, which varies in identity between mealybug hosts. There is strong evidence that across the mealybugs there have been many swapping events, where secondary gamma-proteobacterial participants are introduced and replaced (Thao et al., 2002; Kono et al., 2008). Recent

findings show that the mealybug lineage with the broadest host range, *Planococcus longispinus*, has the most recent secondary symbiont replacement (Husnik and McCutcheon, 2016). It is hypothesised that by retaining flexibility in their associations, mealybug hosts increase their chances of benefiting from novel ecological niches and renewing deleterious gene products, without compromising the critical nutritional functions supplied by their highly degraded primary symbiont *Tremblaya* (Koga and Moran, 2014). The intracellular nature of secondary symbionts within the primary symbiont is critical to the functioning of this model, as the host immune system would otherwise prevent the stable colonisation of so many novel taxa (Husnik and McCutcheon, 2016).

1.1.5 Establishment and maintenance of bacterial communities

Most insects form associations with microorganisms. It is most common for insects to house low-diversity consortia of microbes in the gut (Dillon and Dillon 2004). The insect gut is structurally and nutritionally less complex than that of vertebrates, with a few notable exceptions (e.g. termites), and as such, it houses less complex communities (Douglas, 2011; Engel and Moran, 2013). Like all organisms, invertebrates are constantly exposed to environmental microbes. Some of these microbes are pathogenic, and some are potential symbionts. This poses a problem to the host, as often the recognition molecules are the same for both (Broderick, 2016). In addition, insects lack an adaptive immune response, though evidence for mechanisms of immune memory are now emerging (Nyholm and Graf, 2012).

Immune responses are metabolically costly (Ardia et al., 2012), and constitutive activation detracts from fitness due to trade-offs with reproduction (Sheldon and Verhulst, 1996). Therefore, in order to acquire and maintain beneficial microbes, hosts and their symbionts must evolve mechanisms to reduce the costs of immune activation, whilst maintaining responsiveness to pathogen challenge (Bateman, 1972; Fletcher, 1987; Drew and Yuval, 2000). In addition, the insect gut represents a highly disrupted and unstable environment, particularly in holometabolous insects, where the whole gut organ is shed after every larval instar, completely excised, and reformed

during pupation. Therefore, in order to maintain associations microbes must either be able to survive metamorphosis, or occupy environmental reservoirs and re-inoculate their host after each transition. Despite these challenges, the consistency and simplicity of insect microbial community structures provides evidence that these processes exist, most commonly through specialised morphological structures in the gut (Estes et al., 2009; Stoll et al., 2010).

The insect gut can be a stressful environment, and as such provides constant selection pressures that shape bacterial communities. Host tissues generate specific physiological conditions within the gut to deter pathogens including altered pH, reactive oxygen species (ROS), and immune effectors such as anti-microbial peptides (AMPs) (Appel and Martin, 1990; Lemaitre and Hoffmann, 2007; Buchon et al., 2014). Co-speciation with the host exposes symbionts to a selective environment that can generate resistance to host-specific hostile conditions, which provides a mechanism for hosts to select for beneficial bacteria. The *Burkholderia* symbiont of *Riptortus pedestris* is acquired from the environment at the larval stage of every generation (Kikuchi et al., 2007). A specialised restricted region of the posterior midgut acts to select symbionts from the gut contents and other bacteria by the motility of *Burkholderia* cells and as yet unidentified host factors (Patino-Navarrete et al., 2014). Upon colonization of the midgut crypts, host AMP effectors trigger alterations to the *Burkholderia* cell envelope structure, which stabilizes the exchange of effectors between host and symbiont (Kim et al., 2015). Subsequently, symbionts produce factors such as purines and polyesters, which enable the symbiont to persist within the midgut crypt environment (Kim et al., 2014).

1.1.6 Nutritional interactions

Competition for nutritional resources is an on-going battle between the host and its associated microbes. One evolutionary strategy to attenuate metabolic conflicts is diversifying selection, where each member of the mutualism becomes specialised for the acquisition or utilisation of a particular resource. This is an extremely powerful model in symbioses where partners are intracellular or in very close contact with host cells, enabling metabolic cross

feeding and synergy (Mori et al., 2016). Honeybees house simple communities composed of eight distinct bacterial phylotypes (Martinson et al., 2010). The two most dominant members of these communities are *Gilliamella apicola* and *Snodgrassella alvi*, which show evidence of niche specialization (Engel et al., 2012; Lee et al., 2015). Both have highly reduced genomes and are localized to dense biofilms in the hindgut ileum of the honeybee (Martinson et al., 2012; Kwong et al., 2014). *S. alvi* has lost the ability to import and catabolize carbohydrates, whereas *G. apicola* almost exclusively catabolizes them, providing fermentation products that could be used by *S. alvi* for gluconeogenesis (Kwong et al., 2014). Conversely, *S. alvi* is able to synthesise pyrimidines and amino acids that *G. apicola* cannot. The tight localisation of these bacteria in a highly structured biofilm environment suggests that there may be some metabolic exchange, though this remains to be verified. There is also evidence that the host or its diet may contribute nutrients, as *S. alvi* can survive in the absence of *G. apicola*, implying that metabolic exchange may be dependent on diet or other ecological conditions (Kwong et al., 2014; Lee et al., 2015).

The medfly *Ceratitis capitata* houses a community of diazotrophic and pectinolytic microbes (Behar et al., 2008). Due to their holometabolous lifestyle, the juvenile and adult life stages of medfly have divergent nutritional requirements (Moran, 1994). Eggs are embedded in to rotting fruit through oviposition, and larvae feed on a carbohydrate-rich diet with a high C:N ratio. Adults have a more varied diet of plant exudates, pollen, nectar, bird faeces and bacteria, and have an increased need for nitrogen to facilitate reproduction (Bateman, 1972). Over the course of the life cycle a simple community of *Enterobacteriaceae* is maintained, which changes in structure as the nutritional demands of the host evolve throughout the life cycle. In juveniles, the proportion of pectinolytic bacteria is increased, whereas in adults diazotrophic bacteria are more abundant (Behar et al., 2008). This community of *Enterobacteriaceae* is vertically transmitted from mother to offspring via egg smearing, indicating a stable but adaptive association with its host that is responsive to fluctuations in ecological conditions (Lauzon, 2003).

1.1.7 Nitrogen recycling

Due to the high frequency of herbivory amongst insects and the low-nitrogen content of plant material, many insect bacterial symbionts play a role in supplying and provisioning nitrogen to their hosts (Douglas, 2009). This can be achieved by a number of routes thanks to the diverse metabolic capabilities of bacteria. Some microorganisms are able to increase nitrogen content through fixation: a capability that eludes all eukaryotes (Nardi et al., 2002). Nitrogen fixation has been documented in the gut bacterial communities of termites and beetles (Benemann, 1973; Bridges, 1981), and in the medfly *C. capitata* (Behar et al., 2005). It is also of critical importance to the symbiosis between leaf cutter ants and their fungus gardens, where nitrogen fixation by cultivated fungi contributes significantly to the nitrogen budget of the colony (Pinto-Tomas et al., 2009). However, simply adding more nitrogen in to the system is energetically expensive, relatively rare, and not the only way that bacteria can alter the nitrogen budget of their host.

Like all eukaryotes, insects cannot synthesize all twenty of the amino acids that are required to make proteins, which are commonly obtained from the diet. Howeber, plant sap is nitrogen-poor, with fewer than 20 % of the total amino acids composed of essential amino acids (Douglas, 2006). Despite this, plant sap is widely used as the sole source of nutrition by a number of insects including aphids, scales, whiteflies, psyllids, cicadas and mealybugs (Baumann, 2005). This is thanks to their association with bacterial mutualists, which are able to synthesize the essential amino acids that are lacking (Shigenobu et al., 2000; McCutcheon and Moran, 2010) (Bennett and Moran, 2013). Microorganisms can also recycle host waste nitrogenous compounds. By re-using host metabolic by-products, endosymbionts do not divert resources from the host through competition, and ensure that the maximal nutritional value is extracted from the diet.

Insects excrete nitrogenous waste as uric acid, which can be broken down in to ammonia and urea (Bursell, 1967). It can also be used as a nitrogen store in times of scarcity. In the cockroach *Periplanta americana*, symbiotic *Blattabacterium* are housed in the fat body, where they are able to recycle

ammonia and urea to glutamate. From glutamate they can synthesise all ten of the essential amino acids required by their host (Sabree et al., 2009). Some symbionts are also able to perform the preceding step and break down uric acid to ammonia and urea via uricase enzymes. This is observed in the termite *Reticulitermes flavipes*, where hindgut bacteria provide their host with essential amino acids (Potrikus and Breznak, 1980; 1981), and in *Nilaparvata lugens* by their yeast-like symbionts (Sasaki et al., 1996). The uricase, allantoinase and allantoicase enzymes supplied by *Erwinia*-like symbionts in the midgut caeca of adult *Parastrachia japonensis* shield bugs allow hosts to assimilate amino acids from uric acid (Kashima et al., 2006). In *P. japonensis*, uric acid stores are laid down during nymphal development when food sources are abundant, which is only for around two weeks of the year. These resources are depleted during reproductive diapause, which enables the shield bugs to survive until juvenile food plants become available again. When treated with rifampicin and cleared of their *Erwinia*-like symbionts, shield bugs did not survive reproductive diapause. This was attributed to starvation, as evidence by the reduced concentration of amino acids in the haemolymph (Kashima et al., 2006).

1.1.8 Condition-dependent mutualism

The frequency and spread of facultative symbionts is controlled by their mode of transmission and the prevalence of ecological conditions that favour their host phenotype (Jaenike, 2009). Recent advances in the genomics of hosts and symbionts are beginning to elucidate how genotype-genotype interactions can add to this complexity, including host-symbiont and symbiont-symbiont interactions (Oliver et al., 2006). One parameter that is key in determining phenotype is symbiont titre. High titre leads to higher transmission fidelity and increased phenotypic effects, but must be traded-off with host fitness (Jaenike, 2009; Herren and Lemaitre, 2011). *Hamiltonella defensa* is able to confer resistance to parasitoid wasps in aphids through toxins encoded by APSE-1 bacteriophage integrated into the *Hamiltonella* genome (Oliver et al., 2003; 2009). In *D. melanogaster*, the same phenotype is achieved through replicating *Spiroplasma* cells in the haemolymph, which compete with developing wasp larvae for lipids (Paredes et al., 2016). Both confer a high

fitness benefit in the presence of parasitoids, but are costly to maintain. The high density required to generate resistance to parasites results in reduced longevity (Herren et al., 2014), and reduced lifetime reproductive capacity (Vorburger et al., 2013). In *D. melanogaster*, mechanisms have evolved to delay the onset of this burden beyond the point of peak reproductive capacity. Due to the dependence of *Spiroplasma* on host-supplied lipids, titre is controlled by the nutritional state of the host, which prevents over-proliferation (Herren et al., 2014; Paredes et al., 2016). Similar adaptations have been observed in *D. melanogaster* infected with *Wolbachia*. Chrostek and Teixeira showed that the copy number of *Octomom*, a set of eight genes within the *Wolbachia* genome, is positively correlated with virulence and controls *Wolbachia* titre. Gene copy number can evolve rapidly, and may explain why endosymbiotic organisms that are reliant upon hosts for their own survival are able to rapidly respond to strong selection pressure on host fitness (Chrostek and Teixeira, 2015).

1.1.9 The role of symbionts in mating

In addition to the contributions that bacteria make to fitness through resources such as nutrition or defence against parasites, they can directly influence individual reproductive success by pre- and post- mating isolation. Cytoplasmic Incompatibility (CI) is a common form of post-mating isolation observed amongst insects infected with *Wolbachia*. CI generates high fitness costs in offspring produced from matings between uninfected females and infected males. This drives *Wolbachia* through populations by favouring infected females that are able to vertically transmit the bacterium to their offspring (Perlman et al., 2008).

Microbes also influence pre-mating isolation. Insects largely discriminate between relatives and potential mates by olfactory cues such as pheromones (Martin et al., 2011). In *D. melanogaster*, populations reared on different diets for just one generation displayed assortative mating, preferring to mate with flies reared on the same diet (Sharon et al., 2010). Discrimination between diet types was attributed to the alteration of cuticular hydrocarbons by gut bacteria, though no mechanism for this process has been proposed. Similarly,

Lizé *et al.* 2013 illustrated that manipulation of the gut microbiota alters reproductive investment in *D. melanogaster*. Therefore, either gut microbes are involved in recognition processes or alter physiological cues for reproduction (Lizé et al., 2012). Some insects are unable to copulate or reproduce in the absence of their symbionts. This is the case in plataspid stinkbugs when they are cured of their vertically transmitted gut symbiont *Ca. Ishikawaella capsulata*, which provisions the host with amino acids (Hosokawa et al., 2006). Due to the implicit role of many nutritional symbionts in reaching the physiological state required for reproduction (sexual maturity), it seems logical that variation in bacterial communities may influence mating and reproductive traits (Chippindale et al., 1993).

1.1.10 Practical applications of symbiotic bacteria

Many insects are pests of agricultural crops and act as vectors for diseases of humans, crops, and domestic animals. One strategy that has risen to prominence in recent years for the control of pest insects is to employ the natural associations of insects and microbes in management strategies (Douglas, 2007). Insecticides have been an incredibly valuable in the control of insect pests, but the rapid evolution of resistance alongside the environmental and ethical restrictions of insecticide use validate the need for alternative control strategies. The mosquito *Aedes aegypti* is a globally distributed vector of several human-infective arboviruses and pathogens. *A. aegypti* is not naturally infected with *Wolbachia* (Xi et al., 2005), but introduction of the vertically-transmitted *w*Mel strain from *D. melanogaster* is known to block the transmission of several human pathogens including dengue virus, zika virus, chikungunya virus and *Plasmodium* (Moreira et al., 2009; Hoffmann et al., 2011; Walker et al., 2011). Open field releases of *w*Mel infected *A. aegypti* in an endemic dengue region of northern Australia resulted in the rapid spread of *Wolbachia*, with infection frequency reaching 100 % within a few months of release (Hoffmann et al., 2011). Subsequent recaptures the following year demonstrated that *Wolbachia* infection was stable in the population, and that dengue replication and dissemination to the head were reduced in *w*Mel-infected individuals (Frentiu et al., 2014).

Insect control programs can also employ natural associations between insects and microbes. Paratransgenesis is a technique that exploits existing host-microbial interactions by transgenesis of symbiotic bacteria to express functional proteins that interfere with pathogen survival or transmission (Durvasula et al., 1997). *Sodalis glossinidius* is a vertically transmitted bacterium associated with tsetse flies: the vectors of trypanosomes that infect both humans and animals (International Glossina Genome Initiative, 2014). De Vooght *et al.* successfully transformed *S. glossinidius* with functional anti-trypanosome nanobodies that were able to disperse to multiple tissues within the tsetse fly (De Vooght et al., 2014). Whilst this represents significant progress in the field of paratransgenesis, one considerable challenge to this approach is the delivery of transgenic symbionts to field populations of insects that already have well-established symbiotic microorganisms (Berasategui et al., 2016).

Many control methods exploit the biology of condition-dependent symbionts. These strategies can be risky, as their facultative nature makes them susceptible to variation in environmental conditions (Murdock et al., 2014). Another strategy that has been proposed for the control of agricultural pests such as aphids is to target their obligate symbionts, without which they are unable to survive or reproduce (Douglas, 2007). Concerns over antibiotic resistance instantly prohibit their use for the removal of symbiotic bacteria from pest species. Similarly, catchall approaches to remove bacteria are unlikely to be successful, as many plants also rely on symbiotic bacteria for growth and reproduction. Therefore, control methods in this vein are more likely to require case-specific approaches that incorporate the unique biology of symbiotic interactions.

1.2 The Olive Fruit Fly

1.2.1 Background

Bactrocera oleae (Tephritidae), commonly known as the olive fruit fly, is a multivoltine, holometabolous agricultural pest that lays its eggs in to the mesocarp of olive fruit. Unlike other tephritids, *B. oleae* larvae are able to use both ripening and ripened fruit. This enables them to reach very high population densities and cause significant damage to olive crops despite seasonal variation in the availability of oviposition substrates. *B. oleae* is found throughout the natural range of the genus *Olea* in Europe, the Middle East and Africa, and is also invasive to North and Central America (Tzanakakis, 2003). The olive fruit fly is able to employ all known agricultural cultivars of *Olea* for larval development, including those used for table olives and olive oil (Neuenschwander et al., 1986). They cause significant damage to olive fruit by burrowing through the fruit mesocarp and creating large rot tunnels and exit scars at the fruit surface. While larvae are monophagous on olive fruit mesocarp, and are therefore only ever found in olive groves, adults have a more diverse diet of organic substances. These include insect honeydews, flower nectar, bird droppings, plant exudates, juice from damaged or decaying fruit, and yeasts and bacteria found on or around the leaf phylloplane (Christenson and Foote, 1960; Tsiropoulos, 1977; Sacchetti et al., 2008).

1.2.2 Seasonal population dynamics

B. oleae inhabits temperate climate zones and is subject to seasonal variation in ecological conditions. The polyphagous nature of their diet enables adults to disperse in response to variation in ecological conditions such as heat, humidity, and the availability of oviposition substrates (Bateman, 1972; Fletcher, 1989). In Crete, where the natural populations studied in this thesis were sampled, population density has two main peaks. These occur when temperature, humidity and olive fruit availability are permissive to oviposition: in the autumn, when ripening fruit is at its most abundant, and in the spring, when ripened fruits from the previous year remain on trees (Economopoulos

et al., 1982). Low temperatures and high humidity mean that the majority of the population overwinters as pupae in the soil alongside a small number of adults within or outside of the olive grove (Sacantanis, 1957; Arambourg and Pralavorio, 1970). During the summer months the inverse is true: the majority of the population comprises adults emerged from the spring generation. These adults undergo a period of reproductive diapause between June and July due to a lack of oviposition substrates, high temperature and low humidity (Fletcher et al., 1978). Adults of this generation are primed for peak oviposition rates in the autumn when ripening fruit reaches an adequate size to support larval growth (Fletcher, 1989).

1.2.3 Control methods

The main methods currently employed for *B. oleae* control are bait sprays, cover sprays, and mass trapping, all of which have drawbacks (Haniotakis, 2005). Bait sprays and cover sprays are both insecticide-based strategies that employ organophosphate, pyrethroid and spinosad insecticides. Several populations of *B. oleae* have already evolved resistance to organophosphate insecticides (Vontas et al., 2002), and pyrethroid resistance is well documented in other insect species (Margaritopoulos et al., 2008). These strategies must therefore be used with caution and in concert with other methods.

Mass trapping is a promising strategy that is slow acting but consistent (Haniotakis, 2005). Traps can take many different forms, but usually consist of a vision or odour-based attractant, and a killing mechanism such as sticky adhesive, liquid for drowning, or insecticides (Broumas et al., 2002). This method would ideally be used in tandem with a fast-acting method to reduce population numbers below a threshold value, which could then be monitored and maintained by mass trapping strategies. The employment of *B. oleae* natural enemies for biocontrol programs has previously been suggested as a combinatorial strategy for mass trapping (Daane and Johnson, 2010). Previous biocontrol efforts were unsuccessful due to a mismatch between the natural conditions experienced by parasitoids in their home range, and those experienced in an agricultural setting. However, there has recently been

significant progress and this strategy shows great potential (Daane et al., 2015).

1.2.4 Symbiotic bacteria

All natural populations of *B. oleae* studied to date house the symbiotic bacterium "*Candidatus Erwinia dacicola*". Petri first discovered *Ca.* E. dacicola in 1909 in the digestive tract of *B. oleae*, though its identity remained elusive until 2005 (Petri, 1909; Capuzzo et al., 2005). Natural populations of *B. oleae* house *Ca.* E. dacicola at varying frequencies and relative abundances throughout the life cycle (Petri, 1909; Capuzzo et al., 2005; Estes et al., 2012a; Ben-Yosef et al., 2015). The observation of bacteria on the surface of deposited eggs suggests that egg smearing is the most likely method for vertical transmission. While *Ca.* E. dacicola is known to colonize newly emerged adults, possibly through the infection of midgut progenitor cells during pupation, the mechanisms by which symbionts survive metamorphosis and reformation of the gut have not yet been elucidated (Estes et al., 2012a). It seems unlikely that *Ca.* E. dacicola is re-acquired from the environment by horizontal transmission after each moult, as it cannot replicate outside of the host (Capuzzo et al., 2005; Estes et al., 2009).

Ca. E. dacicola is found throughout the digestive tract, transitioning from an intracellular inhabitant of the larval caeca to a member of extracellular biofilms in adult guts (Estes et al., 2012a). Dense aggregations of *Ca*. E. dacicola form in the oesophageal bulb: a diverticulum of the foregut exclusive to adults that is thought to act as a reservoir for symbionts (Capuzzo et al., 2005). In concert with its shift in tropism between life stages, *Ca.* E. dacicola alters its contribution to host fitness as development progresses. In larvae, when flies are feeding exclusively on olive fruit, *Ca.* E. dacicola facilitates the digestion of organic material loaded with phenolic and potentially toxic compounds (Ben-Yosef et al., 2015). Whilst maintaining a role in nutrition into host adulthood, *Ca.* E. dacicola is thought to transfer its resources towards the recycling of host nitrogenous waste products in adult flies (Ben-Yosef et al., 2010) (Ben-Yosef et al., 2014). Despite evidence of vertical transmission and its contribution to fitness throughout the life cycle, the association between *B.*

oleae and *Ca.* E. dacicola is condition-dependent and non-ubiquitous (Estes et al., 2012a). For these reasons the association more strongly resembles a facultative symbiosis transitioning towards obligacy than one that is already fixed.

In addition to *Ca.* E. dacicola, other bacteria are found in association with *B. oleae* (Estes et al., 2012b). Culture-dependent studies have illustrated that many of the subsidiary bacteria associated with the olive fly are found in the olive grove environment (Tsiropoulos, 1983; Sacchetti et al., 2008; 2013). For this reason, it is assumed that they are allochtonous microbes consumed with the diet that either have rapid turnover in the *B. oleae* gut, or are directly hydrolysed as an alternative food source (Tsiropoulos, 1983; Lloyd et al., 1986; Estes et al., 2012b). The frequency and distribution of transient taxa varies with season, geographic location and population, with the greatest disparity observed between natural and laboratory-reared populations (Estes et al., 2012b).

1.2.5 The Sterile Insect Technique

The Sterile Insect Technique (SIT) is a method that aims to achieve local population eliminations of pest species through the gradual reduction of effective population size. It has been successfully employed against numerous insect species, including several members of the *Tephritidae* (Klassen and Curtis, 2005). In SIT, large numbers of the target insect species are reared, sterilized, and released in to the pest population where they compete for matings with wild insects. Any progeny produced from a sterile-wild mating is non-fertile, which generates a smaller F1 population. The release process is then repeated until the target population has been eliminated or suppressed to an adequate size.

Irradiation is traditionally used to sterilise male insects for release in SIT programs (Economopoulos et al., 1977). Gamma-irradiation damages the gut tissues of mass-reared medfly, which alters the composition of the gut microbiota (Lauzon and Potter, 2012). Fitness deficits attributed to impaired gut function subsequently reduce the success of irradiated males in mating

competitions against wild males (McInnis et al., 1996; Cayol, 1999). Probiotics that harness the attributes of native members of the microbiota have been used to improve mass-rearing methods for SIT-release medfly. Supplementation of the adult diet with *Klebsiella oxytoca*, one of the main components of the wild adult microbiota, rescues this phenotype by reducing the mating latency of lab-reared males, presumably through contribution to nutrition or mating cues (Ami et al., 2009). In addition, the use of a probiotic *Enterobacter sp.* during medfly juvenile development speeds up development time and increases pupal and adult recovery rates, increasing rearing efficiency (Augustinos et al., 2015).

Similar efforts were made to develop an SIT program for olive fruit flies, but olive fruit flies reared in the laboratory were unhealthy due to the sterilisation process, asynchronous in mating period to wild flies, and impossible to rear in a cost and time-effective manner (Economopoulos and Zervas, 1982). Recent advances in our understanding of the importance of microbes in the life cycle of *B. oleae* have reinvigorated interest in SIT for *B. oleae* (Estes et al., 2012b). In addition, novel methods of sterilisation that circumvent the high fitness costs and time-consuming process of gamma-irradiation have been developed in the intervening period (Ant et al., 2012).

1.2.6 Female-Specific Release of Insects with a Dominant Lethal

Female-specific Release of Insects with a Dominant Lethal (fsRIDL) is a method of sterilisation that requires transformation of the target organism with a repressible lethality gene (Thomas, 2000). In the genetically sexing, genetically sterile homozygous OX3097D-Bol strain of *B. oleae* that was developed by Oxitec Ltd. for SIT, repressibility is achieved with the tetracycline-resistance tTav expression system (Ant et al., 2012). This technology exploits alternative splicing between males and females to induce female-specific lethality in the absence of the repressor. Repressibility allows the generation of large numbers of insects for release through conventional rearing methods with the addition of dietary tetracycline. During the release generation insects are reared in the absence of tetracycline, causing female-specific mortality.

Male-specific releases have many advantages. Damage caused by the deposition of eggs by females and subsequent juvenile development in agricultural produce is the most damaging feature of fruit fly pests. By avoiding the release of additional females in to the wild population male-only release programs aim to achieve population suppression without causing further damage to crops. Previous studies have indicated that insects reared in the same environment prefer to mate with each other than with those from a different environment (Sharon et al., 2010). This may be due to ecological mating or kinship cues (Lizé et al., 2012), or to behavioural differences generated by the alteration of selection pressures in a laboratory as opposed to a natural environment (Economopoulos and Zervas, 1982).

1.2.7 Harnessing symbiotic bacteria to improve mass rearing

Though we know that bacteria are important for *B. oleae* fitness in wild populations, we do not understand how they enable larvae to develop in ripening fruit, or adults to reproduce on nitrogen-poor diets. Nor do we know what bacteria stand to gain from occupying the *B. oleae* gut. Artificial diets and sterile laboratory settings cause dramatic shifts in the composition of bacterial communities and significant decreases in *B. oleae* fitness (Economopoulos, 2002; Estes et al., 2012a). Whilst these traits are correlated, no causative link has yet been established between perturbation of the native microbiota and a subsequent reduction in fitness in laboratory flies. Attempts to develop a probiotic diet for *B. oleae* that is able to supplement the laboratory gut microbiota have so far been unsuccessful. Treatment with the naturally associated *Pseudomonas putida*, which is found transiently in *B. oleae* guts and in olive groves, illustrates the complexity of the interaction between *B. oleae* and bacteria (Sacchetti et al., 2013). In order to improve rearing practices, we first need to better understand how bacteria come to be associated with *B. oleae*, how they contribute to *B. oleae* fitness, and how this changes under varying ecological conditions.

1.3 The Tsetse Fly

1.3.1 Background

Tsetse flies are obligate blood feeders found in Sub-Saharan Africa. They transmit African trypanosomes, which are the etiological agents of sleeping sickness in humans and nagana in animals (Aksoy, 2011). Tsetse are viviparous, rearing larvae *in utero* through milk gland secretions and depositing them just hours before pupation. For this reason, tsetse flies are dependent on blood as their sole source of nutrition throughout the life cycle. The genus *Glossina* can be divided in to three ecologically and genotypically distinct subgroups: the *fusca* group, which occupies forests; the *morsitans* group, which occupies savannah environments; and the *palpalis* group, which occupies riverine environments (Leak, 1999; Dyer et al., 2008). Variation in host preference reflects the variation in ecological conditions experienced by each subgroup (Clausen et al., 1998).

1.3.2 Symbiotic bacteria

Tsetse flies vertically transmit three endosymbiotic bacteria. The obligate endosymbiont *Wigglesworthia glossinidia* is housed in specialised bacteriocyte cells in the midgut bacteriome (Reinhardt et al., 1972). It is also found in the tissues of the milk glands, and is vertically transmitted to larvae in milk gland secretions (Aksoy, 1995; Attardo et al., 2008). *Wigglesworthia* has a reduced genome that is 700kb in size and retains several pathways for vitamin biosynthesis, which supplement the fly's restricted diet of vertebrate blood (Akman et al., 2002; Attardo et al., 2008). Females are unable to reproduce in the absence of *Wigglesworthia* (Nogge, 1976), and there is phylogenetic congruence between *Wigglesworthia* and *Glossina*, indicating fidelitous vertical transmission between generations (Chen et al., 1999).

In addition to primary nutritional symbionts, tsetse flies associate with facultative symbionts. *Wolbachia* is localized to the reproductive organs and is found in the somatic tissues of some *Glossina* species (Cheng et al., 2000). The role of *Wolbachia* in tsetse has not been thoroughly investigated, but the

ability to induce cytoplasmic incompatibility in laboratory-reared flies indicates that it is a reproductive parasite (Alam et al., 2011). *Sodalis glossinidius* is a commensal bacterium that is vertically and horizontally transmitted between individuals (Attardo et al., 2008; De Vooght et al., 2015). Maternal vertical transmission is achieved by the same pathway as *Wigglesworthia*: through the milk gland (Attardo et al., 2008). Paternal vertical transmission occurs by sexual transmission of *S. glossinidius* from males to females during mating, and subsequent vertical transmission to intrauterine progeny in milk gland secretions (De Vooght et al., 2015). The genome of *S. glossinidius* reflects this mixed mode of transmission. It lacks congruence with the host phylogeny (Toh et al., 2006) and exhibits strain-level diversity between individuals (Geiger et al., 2006; Farikou et al., 2011), indicating frequent horizontal transmission events. *S. glossinidius* can be cultured *ex vivo* using standard microbiological techniques (Dale and Maudlin, 1999), and though the genome is similar in size to those found in free-living bacteria at approximately 4 Mb, up to 50 % is thought to comprise pseudogenes (Toh et al., 2006). This evidence indicates that *S. glossinidius* is in the early stages of host adaptation, though its role within the host is not well established.

1.3.3 Gut bacteria and vector competence

The gut microbiota has been implicated in susceptibility to trypanosome infection, which is a maternally inherited characteristic in tsetse flies (Maudlin, 1982). *G. m. morsitans* flies with a high infection rate of a *Rickettsia*-like organism (RLO) were more susceptible to midgut trypanosome infection than *G. austeni* flies with a low rate of RLO infection. RLO's are thought to secrete endochitinases that digest midgut chitin to N-acetyl-glucosamine, which subsequently interferes with tsetse trypanocidal lectins, increasing vector competence (Welburn et al., 2009). The identity of this RLO was presumed to be *S. glossinidius* (Dale and Maudlin, 1999), but Dennis et al. recently identified no link between *S. glossinidius* and trypanosome infection rates (Dennis et al., 2014). In addition, the identification of a new *Rickettsia* sp. in *G. m. submorsitans* suggests that the genotypic and phenotypic characteristics of the RLO should be revisited (Mediannikov et al., 2012).

Evidence to date suggests that there is a wealth of bacterial diversity and metabolic complexity within the tsetse fly microbiota that has not yet been discovered. Communities are not restricted to the characterised endosymbionts *Wigglesworthia*, *Sodalis* and *Wolbachia* (Geiger et al., 2009; 2010; 2011; Lindh and Lehane, 2011), but we know little of how these facultative microbes affect host fitness or interact with vertically transmitted members of the microbiota. We also know little of how well characterised vertically transmitted microbes influence each other.

1.4 Objectives

The objectives of this thesis were to describe the structure, composition, and function of bacterial communities associated with two insects of agricultural and epidemiological importance: the olive fruit fly and the tsetse fly. I hope to assess the evolutionary and functional significance of interactions between insects and bacterial consortia, rather than single obligate mutualists. Specific aims were to:

1. Describe the structure and composition of bacterial communities associated with *B. oleae* under varying ecological conditions (Chapter 2).
2. Assess the metabolic complementarity between stable and transient members of the *B. oleae* microbiota, and to determine the contribution of bacteria to host metabolism (Chapter 3).
3. Investigate how perturbation of native bacterial communities affects *B. oleae* health (Chapter 4).
4. Describe the structure and composition of bacterial communities associated with several genotypically and ecologically distinct *Glossina* spp. (Chapter 5).

Chapter 2 : Variation in Bacterial Communities Associated with *Bactrocera oleae*

I designed and conducted the experiment, performed the data analysis, and wrote this chapter.

2.1 Abstract

Background: The olive fruit fly *Bactrocera oleae* is a key pest of olives and imposes a significant financial burden upon olive producers worldwide. As with many insects, the symbiotic bacteria associated with the olive fly contribute to host fitness. The mutualist "*Candidatus Erwinia dacicola*", which plays a role in nutrition, is not ubiquitous in wild populations, and is lost from laboratory populations of *B. oleae* reared on artificial diets. It is not currently known whether other bacteria stably colonize in place of or alongside *Ca.* E. dacicola, or if variation in ecological conditions creates instability in the microbiota that affects host fitness.

Methods: We used 16S rRNA gene amplicon sequencing on the Illumina MiSeq and Pacific Biosciences RS II platforms to evaluate the structure and composition of bacterial communities in two wild-caught and two laboratory-reared cohorts of adult *B. oleae*. The laboratory cohorts were reared with identical methods in two similar laboratory environments, and the wild cohorts were collected from the same geographical location at two different times of year.

Results: Alpha and beta diversity varied between sample cohorts from all populations, and community composition was significantly different between laboratory cohorts. There was significant variation in *Ca. E. dacicola* relative abundance between cohorts. However, this does not seem to be a result of seasonal variation, but of stochastic environmental conditions.

Conclusions: The *B. oleae* microbiota is susceptible to perturbation in both natural and laboratory environments. The causes of perturbation were not investigated, but variation in community composition and structure in both environments suggests that the establishment and maintenance of the *B. oleae* microbiota is sensitive to ecological conditions. This is contrary to many obligate symbionts, which are ubiquitous and maintained independently of stochastic environmental fluctuations.

2.2 Introduction

B. oleae is a globally distributed, economically important pest of olive crops (Daane and Johnson, 2010). Adult females oviposit into olive fruit, where their offspring spend the entirety of their juvenile lives before emerging to pupate in the soil. Larvae feed exclusively on the fruit of the olive, causing significant damage to olive crops. Like many insects, *B. oleae* harbours symbiotic bacterial communities that contribute to host fitness(Dillon and Dillon, 2004). A major component of the *B. oleae* microbiota is the vertically transmitted, non-culture-viable symbiont "*Candidatus Erwinia dacicola*" (Capuzzo et al., 2005; Sacchetti et al., 2008; Estes et al., 2009; Ben-Yosef et al., 2015). *Ca. E. dacicola* is housed in specialized compartments of the gut throughout the life cycle (Estes et al., 2009; 2012a) where it supplements the nutrient-poor diet of adult *B. oleae* by provisioning dietary nitrogen (Ben-Yosef et al., 2014) and enables larvae to develop in ripening fruit by the detoxification of plant secondary metabolites (Ben-Yosef et al., 2015). However, despite its obligate nature, contribution to nitrogen provisioning, and vertical transmission from mother to offspring, *Ca*. E. dacicola infection is not ubiquitous in wild populations (Sacchetti et al., 2008; Estes et al., 2009). Infection frequency ranges from 75-100 % and varies with geographic location (Capuzzo et al., 2005; Kounatidis et al., 2009; Estes et al., 2012a). The factors that influence infection frequency have not yet been elucidated. Additionally, upon introduction of *B. oleae* to the laboratory and transition to an artificial diet, *Ca.* E. dacicola infection is uniformly and rapidly lost in adults. When maintained on olive fruit during larval development, laboratory-reared insects maintain *Ca. E. dacicola* infection throughout development (Hagen, 1966; Estes et al., 2012a). These findings indicate that the interaction between host and symbiont is conditional, and that it is likely mediated by nutritional requirements.

B. oleae experience seasonal fluctuations in dietary composition and ecological conditions throughout their developmental cycle. In Europe, *Olea europaea* flowering occurs in mid to late spring, and fruit grow in size during

the summer, ripen in the autumn, and remain on trees until the following spring (Tzanakakis, 2003). Ripening alters the nutritional and phenolic content of olive fruit (Amiot et al., 1986), and also determines the seasonal availability of oviposition substrates, oviposition cues, and food sources (Fletcher et al., 1978; Kapatos and Fletcher, 1984; Koveos and Tzanakakis, 1990). *Ca.* E. dacicola is essential for the utilization of unripe but not ripe olive fruit by *B. oleae* larvae, due to its role in the detoxification of phenolic compounds, which decrease in concentration during fruit ripening (Fytizas and Tzanakakis, 1966; Hagen, 1966; Ben-Yosef et al., 2015). Similarly, *Ca. E. dacicola* has been shown to enhance female fecundity in a diet-dependent manner (Ben-Yosef et al., 2014). Given that there is variation in the availability of food sources throughout the year (Michelakis and Neuenschwander 1981), and in the microbial reservoirs that are present in the surrounding habitat (Sacchetti et al., 2008), some degree of plasticity in the structure and composition of microbial communities in response to ecological variation is to be expected. This has been observed previously in culture-based studies, where the bacteria recovered from *B. oleae* homogenate varied on a seasonal basis (Tsiropoulos, 1983). In studies where relative abundance was measured, *Ca.* E. dacicola comprised >99 % of the community in larval and >94 % of the community in adult flies (Ben-Yosef et al., 2014; 2015). This corroborates previous findings from culture and molecular-based studies that the bacterial assemblages in wild olive flies are not composed entirely of *Ca.* E. dacicola, but include additional low abundance, and possibly transient, taxa (Tsiropoulos, 1983; Sacchetti et al., 2008; Kounatidis et al., 2009).

Despite disparities in host diet and environmental conditions, bacterial communities associated with species across the *Tephritidae* share a common community structure. Similarly to *B. oleae*, there tends to be one dominant taxon that constitutes more than 80 % of the community. The remainder of the community is comprised of low diversity taxa that fluctuate in frequency and relative abundance (Aharon et al., 2012; Morrow et al., 2015). Many bacterial-*Tephritidae* interactions are nutritional and occur in the gut (Lloyd et al., 1986), and fluctuations have been observed in response to dietary changes such as those that occur during metamorphosis (Andongma et al., 2015). It is

feasible that flexibility in community structure may be maintained as a strategy to mitigate seasonal variation in ecological conditions. This is observed in many insects that house facultative symbionts alongside obligate primary endosymbionts (Feldhaar, 2011). Facultative symbionts provide conditional fitness advantages under specific ecological conditions such as parasitism or heat stress (Oliver et al., 2010). They can also serve as an evolutionary safety net in the event of functional degeneration by primary symbionts (Koga and Moran, 2014). Many endosymbionts have reduced genomes as a result of the population dynamics that they experience during vertical transmission (Moran, 1996). For this reason, ancient partnerships can evolve beyond their peak fitness (Bennett and Moran, 2015). Strategies to avoid such "evolutionary rabbit holes" include symbiont swapping (Koga and Moran, 2014) and the "shopping bag" hypothesis (Larkum et al., 2007), where the evolution of endosymbiosis is a continuous process involving multiple partners (Husnik and McCutcheon, 2016). Both of these strategies require the maintenance of consortia that are able to adapt to ecological parameters, such as is observed in tephritids.

In order to assess how ecological processes, and their absence, might shape the microbial communities associated with *B. oleae*, we used 16S rRNA gene profiling of the V4 region of the 16S rRNA gene to examine bacterial community structure and composition in four cohorts of adult flies. In addition, results were validated with long-read amplicon sequencing of the full-length 16S rRNA gene from a subset of samples. Results confirmed the presence of multiple closely related members of the *Enterobacteriaceae* alongside *Ca.* E. dacicola. In order to assess communities under natural conditions, we collected two populations of wild flies in Crete at different times of year: one in spring and one in autumn. We also examined communities in the absence of dietary selection pressure and natural microbial reservoirs by profiling two cohorts of laboratory strain flies that were reared using standard mass-rearing practices. Our results provide insight in to the flexibility of the *B. oleae* microbiota. Bacterial community structure in laboratory-reared flies resembled that observed in wild *B. oleae* and other Tephritids. Bacteria associated with wild flies displayed seasonal variation in community structure and

composition, and *Ca.* E. dacicola titre varied between cohorts. Alongside *Ca.* E. dacicola, a novel *Tatumella* sp. was detected in eggs and adults from all cohorts studied, including those from Israel. Our findings indicate the susceptibility of the *B. oleae* microbiota to perturbation under varying ecological conditions, and the maintenance of more than one facultatively associated bacterium in wild populations.

2.3 Methods

2.3.1 Insect rearing

Rearing was performed in a laboratory maintained at 25 ^{o}C and 60 % relative humidity. Infected olives were collected and suspended over sterile sand and wandering larvae were allowed to emerge and pupate. Pupae were isolated and stored in plastic cages and upon emergence adults were provided with distilled water and artificial diet (egg yolk, icing sugar and hydrolysed yeast) *ad libitum*. The Argov laboratory strain of *B. oleae* (Ant et al., 2012) was cultured following standard mass-rearing methods (Genc and Nation, 2008), which are outlined in detail in Appendix 1.

2.3.2 Sample cohorts

Two cohorts of Argov strain flies were analysed: Lab1, which was reared at Oxitec Ltd. (Abingdon, Oxfordshire) in 2013, and Lab2, which was reared at the University of Crete (Heraklion, Crete) in 2014. Wild flies were collected from two locations in Crete (Greece): Crete Spring 2013 in Viannos in February, Crete Autumn 2014 in Heraklion from October to December, and Crete Spring 2016 in Heraklion in February. Sample cohort details are listed in Table 2.1. Individual insects were used for all analyses excluding the Lab1 cohort, which included 9 individuals that were dissected in to head, thorax and abdomen prior to DNA extraction. No significant differences were found in

Table 2.1 | Details of the sample cohorts employed in the 16S rRNA gene amplicon study that were sequenced on the Illumina MiSeq platform. *Samples from the study by Ben-Yosef *et al.*, 2015.

Sample Cohort	Origin	Number of Individuals
Lab1	Laboratory	12
Lab2	Laboratory	26
Crete Spring 2013	Wild	21
Crete Autumn 2014	Wild	60
Crete Spring 2016	Wild	22
Israel Wild*	Wild	9
Israel Lab*	Laboratory	6

alpha or beta diversity between the dissected tissues after 16S rRNA gene analysis so were not distinguished in subsequent analyses.

Samples employed for Pacific Biosciences sequencing included some samples that were profiled on Illumina MiSeq, and some new samples that were collected to obtain large amounts of DNA to match the increased input required for PacBio sequencing. Argov Adults were ten adults from the Argov laboratory population that were combined prior to DNA extraction; AMA2 was an individual Argov adult from the Lab1 cohort; Crete Eggs were embryos that were dissected from olives collected in November 2013 and pooled prior to DNA extraction; Crete Adults were ten adults collected in Heraklion, Crete, in November 2013; and CA4 and CA19 were individual adults from the Crete Spring 2013 cohort.

2.3.3 DNA isolation

DNA was extracted using the Qiagen DNeasy Blood and Tissue protocol for Gram-positive bacteria (Qiagen, Manchester, UK) following the manufacturers' instructions. An additional bead-beating step using 3 mm carbide beads (Qiagen, Manchester, UK) in a Qiagen tissue lyzer (Qiagen, Manchester, UK) at 25 Hz for 30 seconds was employed to enhance homogenization of the tissues.

2.3.4 PCR amplification and multiplex sequencing of 16S rRNA genes

All individuals from the Lab1, Lab2, Crete Spring 2013 and Crete Autumn 2014 cohorts were community-profiled by sequencing variable region 4 (V4) of the 16S rRNA gene on an Illumina MiSeq sequencer. Samples indexed with Golay barcodes (Lab1 and Crete Spring 2013 cohorts) were subjected to a total of 50 PCR cycles. An initial round of PCR was performed using universal 16S primers UFPL (5'-AGTTTGATCCTGGCTCAG-3') and URPL (5'-GGTTACCTTGTTACGACTT-3'). PCR reactions were carried out in a volume of 25µl containing 1µl of template DNA, 12.5µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs), 0.5µl of each primer at 10mM concentration and 10.5µl PCR-clean water. Thermal cycling conditions were 98^{o}C for 60 seconds, 25 cycles of 98^{o}C for 60 seconds, 56^{o}C for 45 seconds,

72 ^{o}C for 60 seconds, and a final extension at 72 ^{o}C for 10 minutes. 1 µl of this PCR product was then used as the DNA template in a second round of PCR for the amplification of the V4 region with fusion primers F515 (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GTGCCAGCMGCCGCGGTAA-3') (Caporaso et al., 2012). The 806R reverse primers contained a unique Golay barcode specific to each sample for read de-multiplexing. PCR reactions were carried out in a volume of 25 µl containing 1 µl of template DNA, 12.5 µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs), 1 µl of each primer at 3 mM concentration and 9.5 µl PCR-clean water. Thermal cycling conditions were 30 seconds at 98 ^{o}C, 25 cycles of 98 ^{o}C for 10 seconds, 58 ^{o}C for 30 seconds and 72 ^{o}C for 30 seconds, followed by a final extension of 5 minutes at 72 ^{o}C.

For samples barcoded by dual indexing (Lab2 and Crete Autumn 2014 cohorts), the generation of amplicons for sequencing involved two PCR reactions with a total of 25 cycles of PCR. For the first reaction, primers F515 and 806R (as above) were used to amplify the V4 region. PCR reactions were carried out in a volume of 20 µl containing 5 µl of template DNA at a concentration of 1 ng µl^{-1}, 12.5 µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs), 0.8 µl of each primer at 3 mM concentration and 3.4 µl PCR-clean water. Thermal cycling conditions were 98 °C for 2 minutes, 10 cycles of 98 °C for 20 seconds, 60 °C for 15 seconds, and 70 °C for 30 seconds, with a final extension at 72 °C for 5 minutes. 10.5 µl of this PCR product was cleaned with Agencourt AMPure XP beads (Beckman Coulter Genomics) and used as template for the second PCR reaction. Purified first-round PCR products were combined with 12.5µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs) and 1µl of each barcoding primer including the 8 nucleotide Illumina Nextera index sequences at 3 mM concentration. Thermal cycling conditions were 98 °C for 2 minutes, 15 cycles of 98 °C for 20 seconds, 55 °C for 15 seconds and 72 °C for 40 seconds, with a final extension at 72 °C for 60 seconds.

PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter Genomics) and quantified with the Qubit dsDNA High-Sensitivity

assay (Life Technologies), and an Agilent Bioanalyzer High-Sensitivity DNA chip (Agilent). Samples were pooled at equimolar concentrations and size-selected in a range of 350-450 bp by Pippin-Prep (Sage Science) if required. Sequencing was performed at the University of Liverpool Centre for Genomic Research on an Illumina MiSeq platform with V2 chemistry generating 250 bp paired-end reads.

Pacific Biosciences amplicon profiling of the full-length 16S rRNA gene was performed on a subset of samples as a proof-of-principle experiment to validate the discrimination between closely related members of the *Enterobacteriaceae* using the 300 bp V4 variable region. A 1500 bp region of the 16S rRNA gene was amplified with universal primers 8F (5'–AGAGTTTGATCMTGGCTCAG–3') and 1492R (5'-CCCCTACGGTTACCTTGTTACGAC–3'). Unique Golay barcodes were added to the 8F and 1492R primers for each sample. Libraries were prepared with the Pacific Biosciences 2 kb Library Preparation Kit (Pacific Biosciences) following the manufacturers' instructions. Quality control and pooling were performed as for Illumina MiSeq amplicon libraries, and Dr Margaret Hughes performed sequencing on a Pacific Biosciences RS II sequencer at the University of Liverpool Centre for Genomic Research.

2.3.5 Analysis of 16S rRNA gene Illumina MiSeq amplicon data

Raw sequencing reads were de-multiplexed and converted to FASTQ format using CASAVA version 1.8 (Illumina 2011). Cutadapt version 1.2.1 (Martin, 2011) was used to trim Illumina adapter sequences from FASTQ files. Reads were trimmed if 3 bp or more of the 3' end of a read matched the adapter sequence. Sickle version 1.200 (Joshi and Fass, 2011) was used to trim reads based on quality: any reads with a window quality score of less than 20, or were less than 10 bp long after trimming, were discarded. BayesHammer was used to correct reads based on quality (Nikolenko et al., 2013). Paired-end reads were assembled with a minimum overlap of 50 bp and discarded if outside of the range 200 - 300 bp using PandaSeq (Masella et al., 2012). All subsequent analyses were conducted in QIIME version 1.8.0 (Caporaso et al., 2010b). Sequences were clustered into Operational Taxonomic Units (OTUs)

by *de novo* OTU picking in USEARCH (Edgar, 2010). Chimeras were detected and omitted with UCHIME (Edgar et al., 2011) and the QIIME-compatible version of the SILVA 111 release database (Quast et al., 2013). The most abundant sequence was chosen as the representative for each OTU. Taxonomy was assigned to representative sequences by BLAST (Altschul, 1990) against the SILVA 111 release database (Quast et al., 2013). Representative sequences were aligned against the Greengenes core reference alignment (DeSantis et al., 2006) using PyNAST (Caporaso et al., 2010a).

2.3.6 Statistical analysis of Illumina MiSeq 16S rRNA amplicon data

OTU tables were rarefied to 4400 sequences per sample prior to diversity analyses in order to account for uneven sampling depth. Rarefaction curves were saturated at 4400 sequences per sample, indicating that most of the bacterial diversity was captured at this sampling intensity. All statistical analyses were performed in R version 3.2.3 (R Core Team). Bray-Curtis dissimilarity calculations and NMDS ordination were performed with Phyloseq version 1.14.0 (McMurdie and Holmes, 2013).

2.3.7 Analysis of 16S rRNA gene Pacific Biosciences amplicon data

De-multiplexed Circular Consensus Sequence (CCS) reads were generated by running the Pacific Biosciences Reads of Insert software on raw sequencing data with a minimum pass number of 5 and minimum error rate of 90 % through the SMRT portal. CCS reads were error-corrected with BayesHammer using the option for Pacific Biosciences reads (Nikolenko et al., 2013). Vrevcomp was used to correctly orientate the reads (Hartmann et al., 2011), which were then clustered in to OTUs and processed in an identical manner to Illumina MiSeq amplicons in QIIME version 1.8.0 (Caporaso et al., 2010b), as described previously.

2.3.8 qPCR

qPCR was used to assess the fold ratio increase of *Ca. E. dacicola* 16S rRNA gene copies of Crete Spring 2013 (n=20), Crete Autumn 2014 (n=20), and Crete Spring 2016 (n=22) against Lab2 flies (n=12). Ratios were normalized

to host single-copy alpha elongation factor 1 (EF1α). Primers Edf1 (5'-CTAATACCGCATAACGTCTTCG-3') and EdEnRev (5'-CCACCTACTAGCTAATCCC-3') were used to amplify a 91 bp region of the *Ca. E. dacicola* 16S rRNA gene (Estes *et al.* 2012a). Primers FEF1A (5'-ATCTCGACCAACCGACAAGG-3') and REF1A (5'-ACCAGTTTCAACACGACCGA-3') were used to amplify a 97 bp region of the *B. oleae* EF1α gene. Reactions were performed in a total volume of 20 µl comprising 10 µl SYBR Green PCR Master Mix (Applied Biosystems), 0.6 µl of each 10 µM primer and 2 µl template DNA. The following cycling program was employed: 50 ^{o}C for 2 minutes, 95 ^{o}C for 10 minutes, 40 cycles of 95 ^{o}C for 15 seconds, 52 ^{o}C for 1 minute, and 72 ^{o}C for 10 seconds, followed by a disassociation cycle (95 ^{o}C for 15 seconds, 60 ^{o}C for 30 seconds, and 95 ^{o}C for 15 seconds).

2.4 Results

2.4.1 Community composition and beta diversity in wild and laboratory populations

The bacterial communities associated with laboratory-reared and wild-caught *B. oleae* were investigated with 16S rRNA gene amplicon sequencing. These data were combined with reads generated from one wild and one laboratory population of *B. oleae* collected in Israel for beta diversity analyses (Ben-Yosef et al., 2015). The V4 region of the 16S rRNA gene employed here for bacterial phylotyping does not provide sufficient resolution to identify bacterial taxa below genus level due to the short length of the region and saturation of the phylogeny by deeply diverged lineages. In order to validate the presence of multiple *Enterobacteriaceae* in bacterial communities associated with *B. oleae*, Pacific Biosciences long-read sequencing was used to obtain full-length 16S rRNA amplicons. These reads were added to the SILVA 111 database to improve the taxonomic assignment of V4-region amplicons. Therefore, all V4-region OTUs that were taxonomically assigned to *Ca.* E. dacicola Pacific Biosciences reads are referred to hereafter as such, despite the inability of the region to identify isolates to species level.

Non-Metric Multidimensional Scaling (NMDS) was performed on the Bray-Curtis dissimilarity matrix of OTU occurrence between sample groups and resulted in substantial differentiation between rearing environments and between laboratory and wild-caught flies (Figure 2.1). Wild populations Crete Autumn and Israel Wild, where *Ca.* E. dacicola was the dominant member of the community, formed a tight cluster, as did each of the separate laboratory groups (Figure 2.1,Figure 2.2). The sample group Crete Spring failed to cluster consistently, which upon closer examination was due to the lack of a ‘dominant’ taxon (Figure 2.2). While *Klebsiella* and *Serratia* dominated laboratory communities, *Ca.* E. dacicola, *Tatumella* and *Rhodanobacter* were the highest abundance taxa in wild communities (Figure 2.2). *Ca.* E. dacicola was the most dominant taxon identified in wild flies, with 100 % of Crete Autumn samples (n=60) infected, and 92 % of these infected at a relative

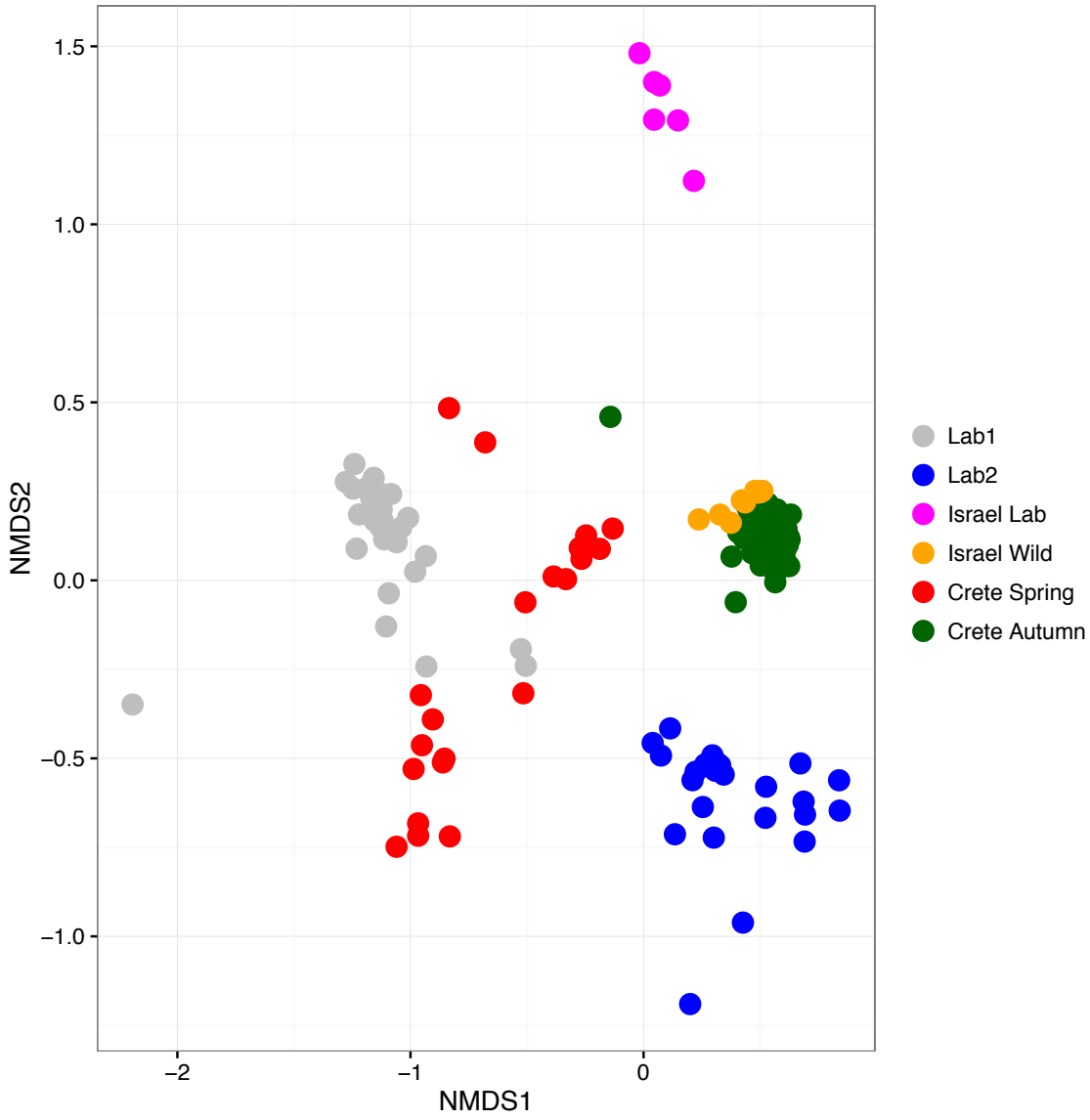


Figure 2.1 | Non-Metric Multidimensional Scaling of a Bray-Curtis dissimilarity matrix of OTUs clustered at 97% similarity. Points represent individual samples and are coloured by sample cohort (Table 2.1).

abundance of 97 % or higher. However, relative abundance and infection frequency were significantly lower in Crete Spring flies, which lacked a dominant taxon. Despite the variation across sample cohorts, all dominant taxa belonged to the family *Enterobacteriaceae*.

Pacific Biosciences sequencing confirmed the results from the Illumina MiSeq 16S rRNA gene V4 region study, validating the presence of novel taxa in the wild population, including *Tatumella* (Figure 2.3). In addition, amplicon

sequencing of eggs that were dissected from olives collected in the autumn (2013) indicates that both *Ca.* E. dacicola and *Tatumella* may be vertically-transmitted during oviposition, or that they are able to colonise the egg within 48 h of deposition.

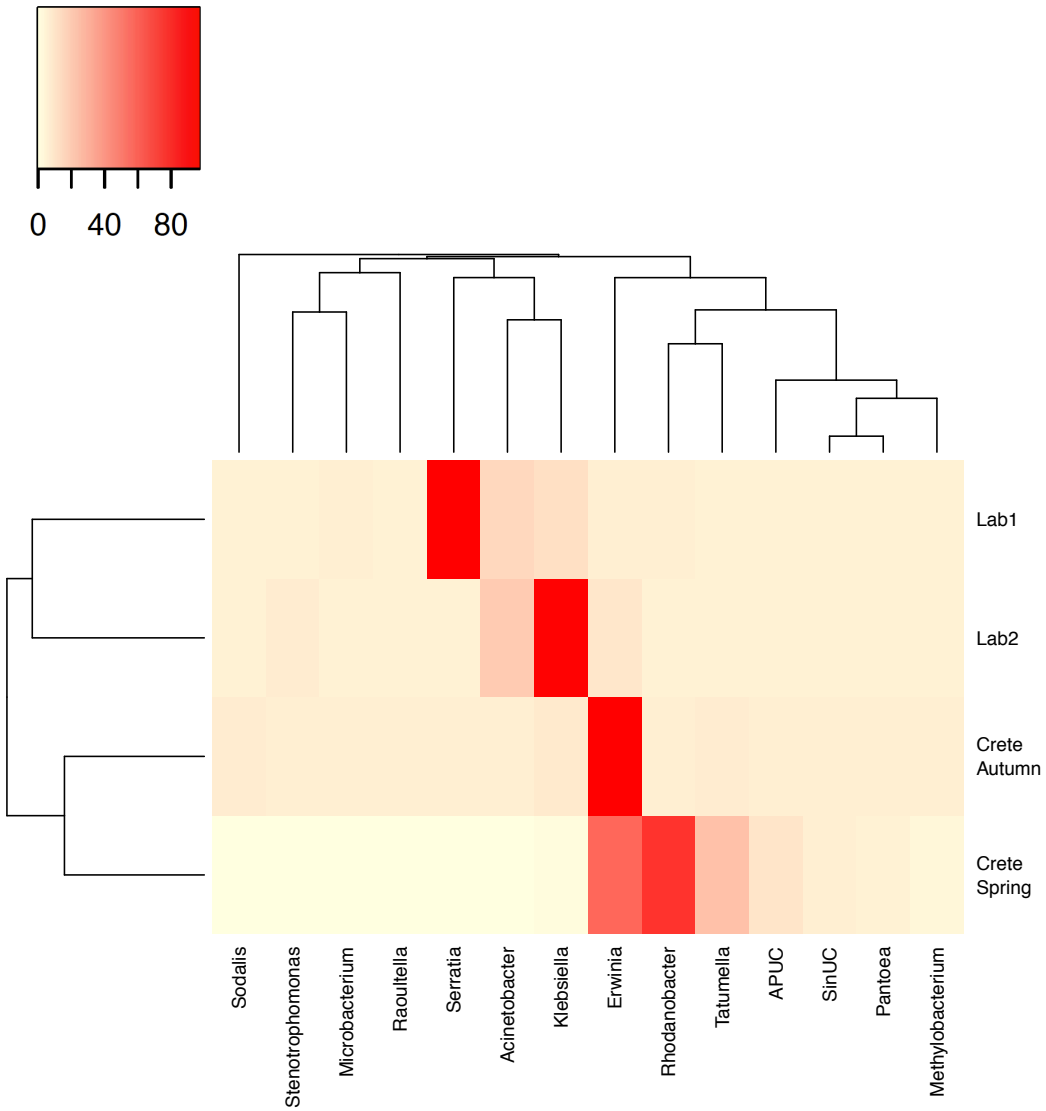


Figure 2.2 | Heatmap showing the relative proportion of 16S rRNA copies of all genera present at >1% relative abundance in any sample cohort as a proportion of the total community in each cohort (Table 2.1). Samples are grouped by hierarchical clustering on both axes based on the Bray-Curtis dissimilarity matrix. The scale bar represents the average relative abundance of 16S rRNA copies of each genus of bacteria (x-axis) per sample cohort (y-axis), as a percentage.

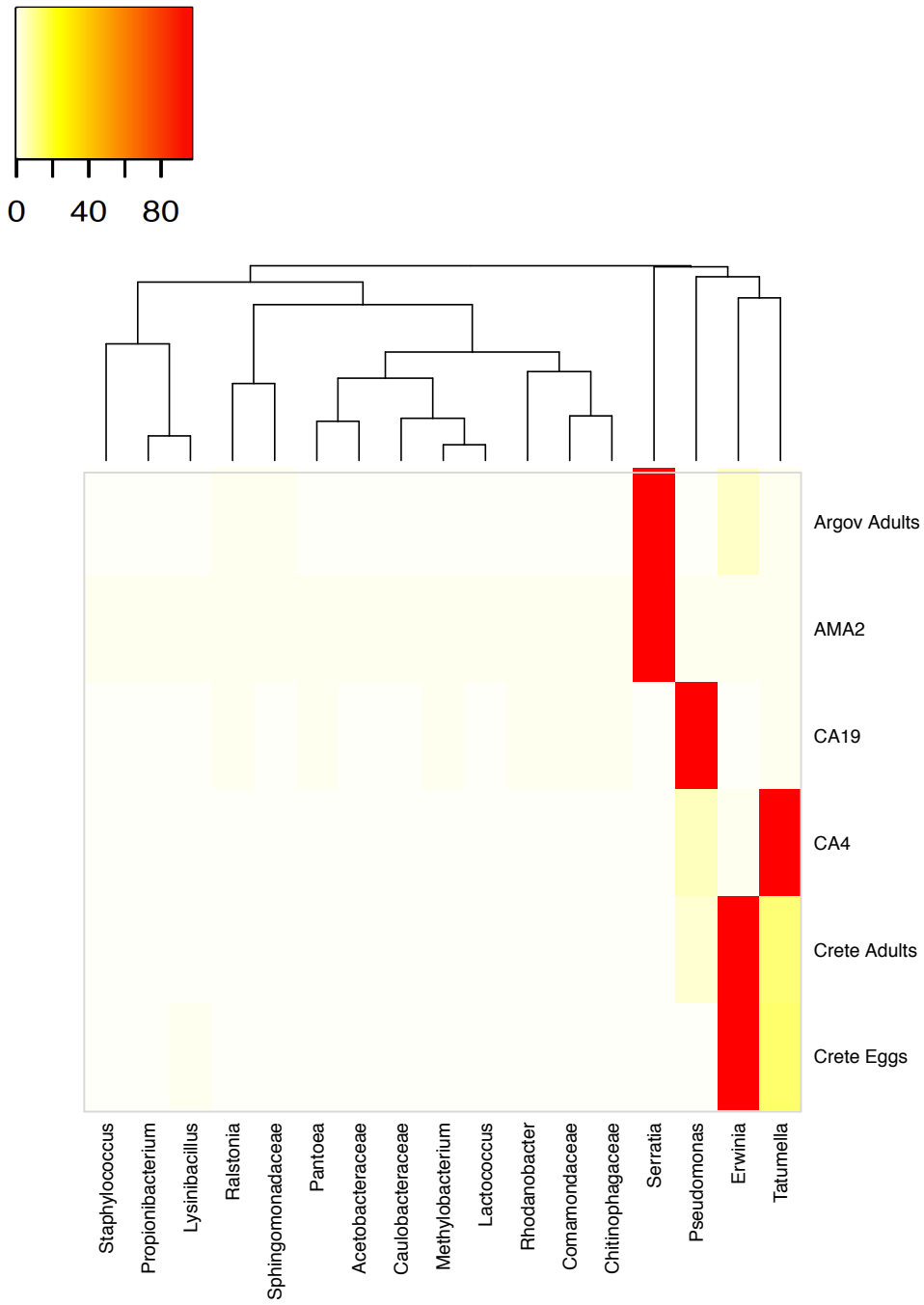


Figure 2.3 | Heatmap showing the relative proportion of 16S rRNA copies of all genera detected by full-length 16S rRNA amplicon profiling using Pacific Biosciences long-read sequencing. Bacterial genera are arranged by hierarchical clustering of the Bray-Curtis dissimilarity matrix. The scale bar represents the average relative abundance of 16S rRNA copy number of each genus of bacteria per sample, as a percentage. Illumina MiSeq profiled samples CA4, CA19, and AMA2 belong to the Crete Spring, Crete Spring, and Lab1 cohorts, respectively. The Crete Adults and Crete Eggs samples were collected in Heraklion in November 2013, and the Argov Adults sample was collected at Oxitec in November 2013. All three samples collected in November 2013 consisted on multiple individuals.

2.4.2 Alpha diversity

Alpha diversity, as measured by the Shannon Index, was significantly different between sample groups (ANOVA, $p < 0.001$, Figure 2.4). Alpha diversity was lowest in Lab1 flies, where the dominant component of the microbiota was *Serratia* (Figure 2.2), and Crete Autumn flies, where *Ca.* E. dacicola comprised ≥97 % relative abundance in the majority of individuals. The sample groups that displayed the highest mean alpha diversity were the Lab2 and Crete Spring flies, which also had the highest within-group variability in alpha diversity (Figure 2.4). A subset of eight flies from the Lab2 cohort displayed significantly higher alpha diversity than the rest of the cohort, and all flies from other cohorts. Interestingly, these samples did not diverge significantly from the rest of the cohort in the NMDS analysis (Figure 2.1).

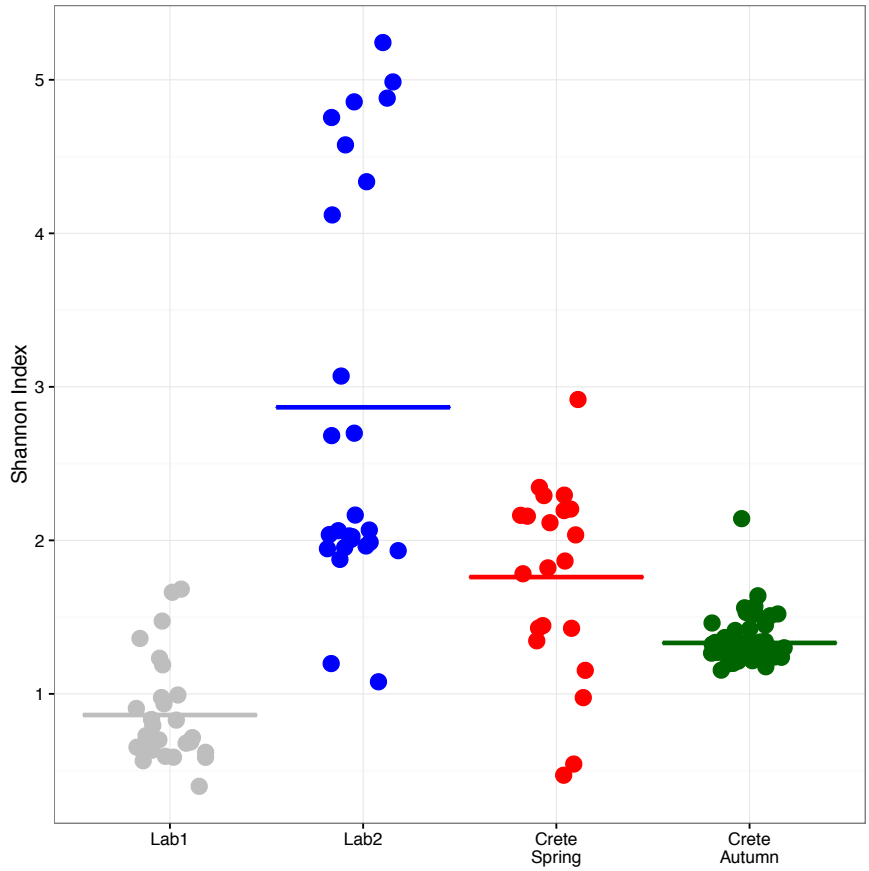


Figure 2.4 | Alpha diversity of the bacterial communities associated with each individual sample as measured by the Shannon index. Individuals are arranged by sample cohort, and bars represent the mean Shannon index of each cohort.

2.4.3 *Ca.* E. dacicola density

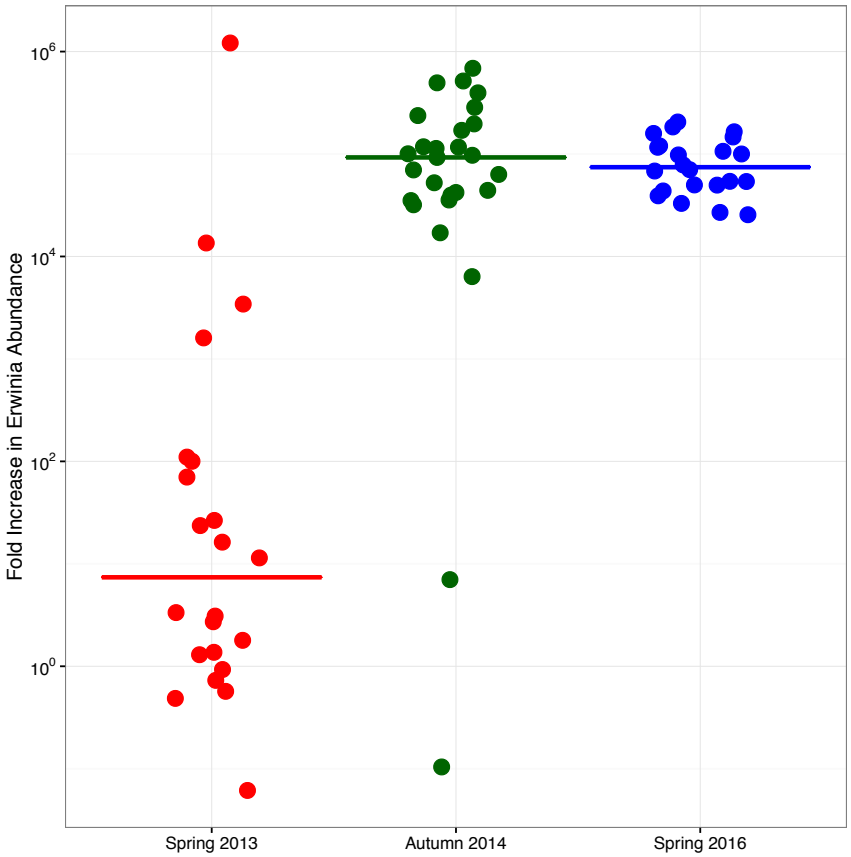


Figure 2.5 | Variation in *Ca. E. dacicola* titre between three wild sample cohorts collected in Crete over the course of three years from 2013 to 2016 as measured by qPCR. Results are represented as the fold-increase of the ratio of *Ca.* E. dacicola 16S rRNA gene copies to host alpha elongation factor copies in comparison to laboratory-reared flies. Points represent individual samples and are arranged by sample cohort, bars represent the median value for each sample cohort.

qPCR was used to measure *Ca.* E. dacicola density, and to validate the low relative abundance of *Ca.* E. dacicola in the Crete Spring population. Results are presented as fold increase in the ratio of *Ca.* E. dacicola 16S rRNA to host alpha elongation factor in comparison to that observed in Lab2 flies. *Ca. E. dacicola* density was significantly higher in Crete Autumn 2014 and Crete Spring 2016 than in Crete Spring 2013 flies (Kruskal-Wallis, $p < 0.001$, Figure 2.5). However, there was one Crete Autumn 2014 individual (C3M) with a low qPCR ratio. 16S rRNA gene amplicon profiling indicated that *Ca.* E. dacicola relative abundance was just 2.4% in this sample, and that the dominant member of the community was a bacterium with a 16S rRNA gene V4 region that is highly phylogenetically similar to that of *Sodalis glossinidius*. The

Sodalis-like bacterium was present at 93 % relative abundance in this individual. NMDS analysis clearly differentiates this sample from the main Crete Autumn 2014 cluster (Figure 2.1), as it was the only individual in this study infected with the *Sodalis*-like bacterium (next-highest relative abundance was 0.07 %). Crete Spring 2013 flies exhibited a wider distribution of *Ca.* E. dacicola densities than the other sample groups assayed (Figure 2.5). One sample from the Crete Spring 2013 cohort had a *Ca.* E. dacicola relative abundance of 97.3 % by amplicon profiling and a qPCR ratio comparable to those observed in Crete Autumn 2014 and Crete Spring 2016 flies, indicating that the reduced titre of *Ca.* E. dacicola in Crete Spring 2013 samples is unlikely to be an artefact of sample collection or methodology.

2.5 Discussion

The structure and composition of bacterial communities associated with *B. oleae* varied between cohorts of adult flies. The Crete Autumn cohort was the only sample group dominated by the nutritional symbiont *Ca.* E. dacicola. *Ca.* E. dacicola infection frequency was previously recorded at 75 % in a wild population of flies from Crete (Estes et al., 2012a). Contrary to these findings, all wild flies were infected with *Ca.* E. dacicola, though at highly variable relative abundance. This was validated with qPCR, which showed high variability in symbiont titre between wild cohorts. 16S rRNA gene amplicon profiling identified three bacterial genera from the *Enterobacteriaceae* in addition to *Ca.* E. dacicola that at high relative abundance: *Sodalis*, *Tatumella* and *Rhodanobacter*. *Sodalis* is a diverse host-associated bacterium that is able to form symbiotic associations with insects from a broad phylogenetic range (Snyder et al., 2011). *Tatumella* has been characterised as a symbiont of *Drosophila suzukii*, and is posited to contribute to nutrition (Chandler et al., 2014). *Rhodanobacter* is a common contaminant of DNA extraction kits and PCR reagents (Salter et al., 2014), and may be at increased abundance in Crete Spring 2013 due to a reduction in total bacterial titre in addition to *Ca. E. dacicola* titre, though this was not tested here. The variation in community composition observed in wild flies during this study indicates that the interaction between *Ca.* E. dacicola and *B. oleae* is not yet stabilised. 16S rRNA gene amplicon profiling has identified that there are other potential mutualists circulating in the system, and that they can be present at high relative abundance in both eggs and adults. However, it must be noted that taxon relative abundances were not corrected for variation in 16S rRNA copy number between bacterial isolates, which is known to skew reported community diversity and structure in 16S studies (Kembel et al., 2012), and therefore are only estimates of the relative proportions of the taxa present.

Many insect symbionts (including *Ca.* E. dacicola) originate from the *Enterobacteriaceae* (Husník et al., 2011), and have been able to form symbioses due to their ability to provide novel functions to their host.

Associations between organisms with distinct metabolic capabilities facilitate rapid adaptation to novel ecological conditions (Moran, 2007; Sudakaran et al., 2015). Tephritids consistently affiliate with low diversity consortia of *Enterobacteriaceae*, many of which contribute to nutrition (Lloyd et al., 1986; Aharon et al., 2012; Morrow et al., 2015), and therefore have a high number of potentially mutualistic partners. It is possible that the process of symbiont selection has not yet advanced to a state of fixation between *B. oleae* and *Ca.* E. dacicola. This would explain the instability observed in the structure and composition of the microbiota, and the apparent susceptibility to variation in ecological conditions. It is also possible that variation in the quality and composition of the diet in both juveniles and adults prevents the requirement for a potentially costly obligate symbiont, and instead favours the maintenance of non-obligate consortia, as is observed in medfly (Behar et al., 2008).

In comparison to the more complex communities found in vertebrates, insect-associated bacterial communities are simple and highly specialised (Douglas, 2011). In vertebrates, decreasing microbial diversity has been correlated with ill health (Donaldson et al., 2015). Due to the naturally low diversity of the invertebrate microbiota, the opposite may be true: increasing diversity signals perturbation. One of the aims of this study was to assess how community diversity is affected by the alteration of microbial reservoirs by mass-rearing practices and ecological conditions that vary seasonally. Previous studies suggest that *B. oleae* houses a low diversity bacterial community that is dominated by one bacterium: *Ca.* E. dacicola (Ben-Yosef et al., 2014; 2015). Whilst the Crete Autumn sample cohort recapitulated this condition, the other three cohorts deviated. The laboratory environment alters the ecological processes that govern bacterial community establishment and maintenance in *B. oleae*. The results of this study consolidate this finding, and also indicate that similar perturbations to community structure and composition can occur in wild flies. The bacterial communities associated with Crete Spring flies displayed reduced titre of the symbiont *Ca.* E. dacicola and, probably as a result, increased variability in composition and alpha diversity in comparison to *Ca.* E. dacicola-dominated flies collected from a similar location. The

ecological processes that drive such perturbations were not investigated over the course of this study, but warrant investigation as potential drivers of stability in simple bacterial communities.

Chapter 3 : Genomic analysis suggests that gut bacteria may contribute to nitrogen recycling in *Bactrocera oleae*

I designed and conducted the experiment, performed the data analysis, and wrote this chapter. An additional *Ca.* E. dacicola genome assembly was provided by Michael Ben-Yosef, Zohar Pasternak, and Edouard Jurkevitch, David Starns assisted with cell sorting and whole genome amplification of *Ca.* E. dacicola isolates, and Anastasia Gioti and John Vontas provided metagenomic shotgun data from *B. oleae* larvae.

3.1 Abstract

Background: "*Candidatus Erwinia dacicola"* is a vertically transmitted mutualistic bacterium that plays a role in nitrogen provisioning throughout the life cycle of the agricultural pest *Bactrocera oleae*. *Tatumella* TA1 is a novel bacterium present at high frequency and low relative abundance throughout the life cycle of *B. oleae* from Crete and Israel, and is hypothesised to be a facultative symbiont. Genome sequences for *Tatumella* TA1, *Ca.* E. dacicola, and two allochtonous members of the gut microbiota were generated to assess the potential roles of bacteria in host fitness.

Methods: The *Ca.* E. dacicola genome was assembled by a hybrid approach, employing metagenomic and single-cell data to generate the ErwSC composite assembly. Draft genomes were generated for transient members of the microbiota *Chryseobacterium* CBo1 and *Stenotrophomonas* SBo1, and *Tatumella* TA1. Genome annotation and phylogenetic analysis of orthologous proteins were conducted to assess the functional repertoire of the community and the phylogenetic origins of *Ca.* E. dacicola and *Tatumella* TA1.

Results: The reduced genome of *Ca*. E. dacicola (2.1 Mb) reveals its phylogenetic origins in the genus *Erwinia*, and the acquisition of a urease operon by horizontal gene transfer. The genome is enriched in factors for amino acid biosynthesis and host nitrogen provisioning, and depleted in those required for carbohydrate metabolism. The *Tatumella* TA1 genome also encodes genes required for amino acid biosynthesis and the metabolism of host waste nitrogenous products. In addition, both encode operons required for the tolerance of physiological conditions encountered in the insect gut such as oxidative stress, iron limitation, and microbe-microbe interactions.

Conclusions: Genome reduction in *Ca*. E. dacicola suggests that it is an obligate symbiont of *B. oleae*. Both *Ca.* E. dacicola and *Tatumella* TA1 encode genes that may be involved in host amino acid and nitrogen provisioning, and functional analysis suggests that there may be stabilising interactions and metabolic cross feeding between the host and multiple members of the microbiota.

3.2 Introduction

Many insects form persistent associations with microbes, which have allowed them to rapidly adapt to novel biotic and abiotic conditions (Feldhaar, 2011; Sudakaran et al., 2015). First discovered by Petri in 1909, the bacterial symbionts of *Bactrocera oleae* (Tephritidae) have been studied for over a century (Petri, 1909). "*Candidatus Erwinia dacicola"* resides in specialised compartments of the digestive tract throughout the lifecycle, and is transmitted to the next generation by egg smearing (Estes et al., 2009). The native microbiota, 75 % or more of which is composed of *Ca*. E. dacicola in all life stages, enhances *B. oleae* fitness by contributing to nutrition (Estes et al., 2012a; Ben-Yosef et al., 2015). Microbes enable larvae to develop in ripening fruit that is abundant in phenolics and toxic compounds (Hagen, 1966; Ben-Yosef et al., 2015), and adults to subsist and reproduce on a nitrogen-poor diet (Ben-Yosef et al., 2010; 2014). The phenotypes that result from symbiotic associations are dependent upon many factors, including genotype-genotype interactions and variation in environmental conditions (Russell and Moran, 2006).

In *B. oleae*, symbiont contribution to host fitness is diet-dependent and only observed in nutrient-poor conditions. *Ca.* E. dacicola is thought to augment host nutrition through nitrogen provisioning, which is critical for both sexes during reproduction (Drew and Yuval, 2000). When fed on a defined diet of sugar and non-essential amino acids, females cured of their microbiota were significantly less fecund than their symbiotic counterparts (Ben-Yosef et al., 2010; 2014). *B. oleae* are holometabolous, which permits the divergence of ecological strategies between juveniles and adults of the same species. For this reason, it is feasible that the host has different functional requirements of the microbiota, and vice versa, between life stages (Moran, 1994). Larvae are able to develop in ripening fruit: a feature that is uncommon amongst fruit flies, and which is dependent upon the presence of the native microbiota (Hagen, 1966). This trait is attributed to supplementation of the homogenous larval diet of olive flesh by resident symbionts, which are able to metabolize

secondary compounds and synthesise essential amino acids that are lacking, particularly lysine (Ben-Yosef et al., 2015). However, the mechanisms by which the microbiota, in particular *Ca.* E. dacicola, augment the diet and contribute to nitrogen provisioning have not yet been elucidated.

Associations between organisms can significantly alter their evolutionary trajectories. Host-associated bacteria display distinctive patterns of genome evolution due to commonalities in transmission and selection pressures. These patterns are enhanced by the evolution of dependency, and are powerful diagnostic features when examining symbiotic interactions. Obligate, intracellular bacteria often experience degenerative evolution as a result of population dynamics and mutational bias inherent in bacterial genomes (McCutcheon and Moran, 2012). Endosymbionts that have been evolving with a host for long periods of time are transferred in small numbers from one generation to the next, which acts to reduce the effective population size and increase genetic homogeneity within populations. These features prevent purifying selection from purging deleterious mutations that accumulate naturally, and results in accelerated genetic drift (Moran, 1996). Mutational bias towards deletion and G-C to A-T transitions results in endosymbiont genomes that are significantly reduced in size, are AT-rich, and have elevated rates of protein evolution (Lambert and Moran, 1998; Fares et al., 2002; Nilsson et al., 2005).

Endosymbionts also experience a different phenotypic selection environment to free-living bacteria. Host homeostasis ensures that, whether bacteria live extracellularly in the gut or in specialised cells, physiological conditions remain relatively constant. The host represents a highly regulated, nutrient-rich environment, which demands a specific set of metabolic and physiological capabilities. The main challenges posed to endosymbiotic bacteria are to evade the host immune system, and establish and maintain nutrient exchange whilst avoiding competition that could harm fitness (Schwartzman and Ruby, 2016). The retained genes in small symbiont genomes reflect these priorities. Sap-sucking insects house symbionts with the most dramatically reduced genomes, all of which retain biosynthesis pathways for essential amino acids

required by the host (McCutcheon and Moran, 2010). Many endosymbionts lose genes that encode proteins involved in cell envelope formation. This adaptation is thought to prevent damaging constitutive activation of the host immune response, and avoid the synthesis of redundant proteins that are not required in the presence of the host cell envelope (Shigenobu et al., 2000; Akman et al., 2002; Gil et al., 2003). Examining the structure, composition, and functional complexity of symbiont genomes provides detailed insight in to the nature and evolutionary history of interactions, and to identify convergent traits between diverse host-microbe interactions.

We aimed to elucidate the nature of the interaction between *B. oleae* and its native microbiota by studying the genomes of two permanently associated and two transiently associated bacteria. Draft genomes were generated for taxa identified as *Ca.* E. dacicola and *Tatumella* TA1, which are putative symbionts of *B. oleae*, and *Chryseobacterium* CBo1 and *Stenotrophomonas* SBo1, which are allochtonous members of the *B. oleae* gut microbiota. Genomes were annotated and the phylogenetic origins of *Ca.* E. dacicola and *Tatumella* TA1 were determined by ortholog analysis. There is an analogous interaction in the Western Flower Thrip (WFT), where an *Erwinia* (BFo1) and *Tatumella* (BFo2) symbiont co-reside, and constitute the most similar phylogenetic relatives of *Ca.* E. dacicola and *Tatumella* TA1 (Facey et al., 2015). Our findings indicate that both *Ca.* E. dacicola and *Tatumella* TA1 show evidence of host adaptation, and that they may play a role in nitrogen provisioning and amino acid biosynthesis through complementary metabolic pathways. Our findings correlate with previous experimental evidence of the contribution of the *B. oleae* microbiota to host fitness. However, we hypothesise that the interaction between *B. oleae* and its microbiota is not limited to single-species reciprocity, and that bacterial communities may be able to dynamically respond to variation in nutrient availability.

3.3 Materials and Methods

3.3.1 Insect material

Multiple populations of wild *B. oleae* were sampled from olive trees in the grounds of the University of Crete (Heraklion, Greece) between October and November 2013 to 2015 by collecting infested olives and dissecting first and second instar larvae from the fruit or allowing third instar larvae to emerge and pupate in sterile sand. Adults were reared on conventional artificial diet (Appendix 1).

3.3.2 Isolation of culture-viable bacteria

Adult *B. oleae* were surface-sterilized in 70 % ethanol and rinsed twice in distilled water prior to dissection in sterile PBS and mechanical lysis of gut tissues with a plastic pestle. Microbiological conditions use to grow culture-viable taxa are listed in Table 3.1.

Table 3.1 | Insect material and microbiological culture conditions used to isolate low-abundance bacterial taxa from natural populations of *B. oleae*.

Tissue	Age (days)	Microbiological Conditions	Incubation Time	Isolate Name	Genus (16S rRNA)
Gut	10	BHI agar, aerobic	72 h	CBo1	*Chryseobacterium*
Gut	10	BHI agar, aerobic	48 h	SBo1	*Stenotrophomonas*
Whole	2	Columbia + 10% DHB	48 h	TA1	*Tatumella*

3.3.3 Identification of isolates

Single colonies were picked with a 1 µl sterile loop and placed in to 10 µl PCR-clean water in a PCR tube. Tubes were incubated at 95 ^{o}C for 5 minutes to lyse cells and isolate DNA. A 1500 bp region of the 16S rRNA gene was amplified with universal primers 8F (5'–AGAGTTTGATCMTGGCTCAG–3') and 1492R (5'-CCCCTACGGTTACCTTGTTACGAC–3'). Reactions were performed in a total volume of 25 µl comprising 12.5 µl MyTaq Red (Bioline), 0.5 µl of each 10 µM primer stock, 1 µl template gDNA, and 10.5 µl PCR-clean water. Thermal cycling conditions were 95 ^{o}C for 5 minutes, 30 cycles of

95 °C for 30 seconds, 56 °C for 45 seconds, 72 °C for 90 seconds, and a final extension at 72 °C for 7 minutes. PCR products were Sanger sequenced with forward primer 8F by GATC (GATC, Cologne).

3.3.4 DNA extraction

Single colonies were inoculated into BHI broth or nutrient agar, incubated at 25 °C, and grown to an OD600 of 0.3. Cultures were pelleted by centrifugation at 6000 x g for 6 minutes, the supernatant removed, and cells were re-suspended in DNA elution buffer. DNA was extracted using the Zymo Quick DNA Universal Kit (Zymo) following the manufacturers' instructions for biological fluids and cells with the following amendment to the protocol: samples were incubated with proteinase K at 55 °C for 30 minutes rather than 10 minutes. TA1 DNA was cleaned up with Ampure beads (Agencourt) at a 1:1 ratio and stored at 4 °C until library preparation. CBo1 and SBo1 DNA were stored at -20 °C.

3.3.5 CBo1 and SBo1 library preparation and sequencing

400 ng of gDNA was sheared with a Bioruptor pico (Diagenode) following the manufacturers' instructions to generate 500 bp inserts. Libraries were prepared with the NEBNext Ultra DNA library preparation kit (New England Biolabs) following the manufacturers' instructions. Adapters were diluted 1:10 to compensate for lower input DNA quantity. Libraries were validated with an Agilent bioanalyzer HS chip (Agilent) and the Qubit fluorometer HS kit (Life Technologies) and were sequenced on an Illumina MiSeq sequencer with paired-end 250 bp reads at the Centre for Genomic Research, University of Liverpool.

3.3.6 TA1 library preparation and sequencing

DNA was sheared to 10 kb using Covaris G-tubes following the manufacturers' guidelines. The sample was purified with Ampure beads (Agencourt) at a 1:1 ratio and library preparation was performed with the SMRTbell library preparation kit (Pacific Biosciences) following the manufacturers' instructions. Resulting libraries were purified with Ampure

beads (Agencourt) at a 1:1 ratio as before. The Qubit dsDNA HS assay (Life Technologies) was used to quantify the library and the average fragment size was determined using the Agilent Bioanalyser HS assay (Agilent). Size selection was performed with the Blue Pippin Prep (Sage) using a 0.75 % agarose cassette and the S1 marker. The final SMRT bell was purified and quantified as previously. The SMRTbell library was annealed to sequencing primers at values predetermined by the Binding Calculator (Pacific Biosciences) and sequencing was performed on two SMRT cells using 360-minute movie times. Dr Margaret Hughes performed Pacific Biosciences sequencing and library preparation of TA1 isolates at the University of Liverpool Centre for Genomic Research on a Pacific Biosciences RS II sequencer.

3.3.7 Isolation of *Ca.* E. dacicola gDNA for whole genome sequencing

Fluorescence Activated Cell Sorting (FACS) was used to isolate individual *Ca. E. dacicola* cells. Guts from ten 16-17 day-old *B. oleae* adults (five males and five females) that emerged from infested olives collected in Heraklion (Crete, Greece) in November 2015 were dissected into 500 µl PBS. Gut tissue was homogenized with a plastic pestle under aseptic conditions. Lysate was pelleted at 4 ^{o}C by centrifugation at 10,000 x g for 5 minutes. The supernatant was removed and the lysate re-suspended in 500 µl PBS. The re-suspended lysate was filtered through a 5 µm filter (CellTrics) by centrifugation at 2000 x g for 1 minute to remove large sections of tissue and aggregates. A 100 µl aliquot of the sample was reserved as a control and the remaining 400 µl was stained with Cell Tracker deep red at a final concentration of 3 µM. The stained sample was incubated at room temperature for 90 minutes in the dark. Following staining, the sample was centrifuged at 4 ^{o}C for 5 minutes at 10,000 x g, the supernatant was removed, and the sample was re-suspended in 400 µl PBS. Samples were kept on ice until sorting. Single cells were sorted into a 96-well plate using a Sony SH800 cell sorter (Sony). All PBS employed for sample preparation and cell sorting was filtered through a 0.22 µm filter prior to use. *Ca.* E. dacicola DNA was also isolated on solid microbiological media, though there were no visible colonies. Guts from five adults were dissected

into 100 µl PBS sterilized with a 0.22 µm filter. Gut tissue was homogenized with a sterile plastic pestle under aseptic conditions and the liquid portion of the sample was pipetted on to BHI agar and spread with a sterile glass spreader. Plates were incubated at 25 ^{o}C for several days. DNA was isolated from the surface of the agar media with a sterile 1 µl loop, transferred to a PCR tube containing 10 µl PCR-clean water and incubated at 95 ^{o}C for 5 minutes to isolate DNA. DNA from a total of 4 single cells and 4 boil preparations from solid microbiological media was whole genome amplified using the QIAGEN REPLI-g Single Cell Kit (Qiagen) following the manufacturers' instructions. All reagents were sterilised by UV irradiation for 30 minutes prior to use.

3.3.8 *Ca.* E. dacicola DNA library preparation and sequencing

All isolates were checked post-MDA by PCR amplification of the 16S rRNA gene and Sanger sequencing as above. Isolates were confirmed by BLAST against the NCBI nt database. 400 ng of DNA per sample was sheared with a Bioruptor pico (Diagenode) following the manufacturers' instructions to generate a 500 bp insert size. Libraries were prepared with the NEBNext Ultra DNA library preparation kit (New England Biolabs) following the manufacturers' instructions. Adapters were diluted 1:10 to compensate for lower input DNA quantity. Libraries were validated with an Agilent bioanalyzer HS chip (Agilent) and the Qubit fluorometer HS kit (Life Technologies). Libraries were sequenced on an Illumina MiSeq with 250 bp paired-end reads at the Centre for Genomic Research, University of Liverpool.

Metagenomic shotgun and mate-pair (2-6 kb) libraries were prepared from gastric caeca dissected from third-instar larvae isolated in Israel from unripe olives and from a mixture of ripe and unripe olives, respectively. DNA for the shotgun library was extracted using an adapted CTAB method (Xin and Chen, 2012) with additional bead beating and lysozyme digestion, and the library was prepared with the Ovation Rapid DR Multiplex System (NuGen). The mate-pair library was prepared with the gel-free NexteraMate protocol from DNA extracted with the Chemagic DNA Bacteria Kit (Chemagen). Both

libraries were prepared and sequenced on an Illumina MiSeq sequencer by LGC Genomics GmbH.

3.3.9 TA1 draft genome assembly from Pacific Biosciences sequencing data

365,445 sub-reads with a mean length of 6,926 bp were assembled with the Hierarchical Genome Assembly Process (HGAP) workflow (Chin et al., 2013). The HGAP pipeline comprises pre-assembly error-correction of sub-reads based on read length and quality, assembly with Celera, and assembly polishing with Quiver.

3.3.10 CBo1, SBo1, and *Ca.* E. dacicola draft genome assembly

Raw sequencing reads were de-multiplexed and converted to FASTQ format with CASAVA version 1.8 (Illumina 2011). Illumina adapter sequences were trimmed from raw reads with Cutadapt version 1.2.1 if 3bp or more of the 3' end of a read matched the adapter sequence (Martin, 2011). Reads were quality trimmed with Sickle version 1.200 (Joshi and Fass, 2011). Any reads with a window quality score of less than 20, or which were less than 10bp long after trimming, were discarded. Paired-end reads were quality checked with fastqc prior to assembly (Andrews, 2010). Libraries were assembled with SPAdes version 3.7.1 (Nurk et al., 2013) in single cell mode with kmer sizes 33 to 127. Reads were error-corrected prior to assembly with BayesHammer as implemented in SPAdes. All eight *Ca.* E. dacicola libraries were assembled separately and in combination to assess individual library quality. Contigs shorter than 500 bp were removed from the final assemblies.

3.3.11 ErwSC hybrid draft genome assembly

Metagenomic shotgun data and eight single cell and boil-preparation libraries for *Ca.* E. dacicola were combined to generate a hybrid genome assembly referred to hereafter as ErwSC. Reads were error-checked and assembled with SPAdes version 3.7.1 as described previously. Due to the presence of multiple *Enterobacteriaceae*, and other bacteria, in the larval gut, the hybrid assembly was assessed and filtered for auxiliary taxa. Using Blobology

(Kumar et al., 2013), contigs identified as belonging to other organisms based on coverage and GC content were excluded, and the 12,519,932 reads that mapped back to putative *Ca.* E. dacicola contigs were extracted and reassembled with SPAdes. Contigs shorter than 500 bp were removed from the final assembly.

3.3.12 Genome annotation and assembly statistics

Coverage was assessed by mapping the raw reads to the draft assembly with Bowtie2 (Langmead and Salzberg, 2012). Taxonomy was assigned to contigs with BLAST, and GC content was calculated with the Blobology package. All non-target contigs were filtered from the assembly based on GC-content, coverage, and taxonomy and TAGC plots were drawn in R version 3.2.3 (R Core Team, 2015). Assembly statistics were calculated using custom perl scripts and Qualimap version 2.2 (Okonechnikov et al., 2016). Assemblies were annotated with PROKKA version 1.5.2 (Seemann, 2014), and completeness was assessed following the method in Rinke et al. (2013). Completeness values refer to the proportion of genes present in the assembly compared to the expected number in free-living bacteria. COG functional categories were assigned by genome annotation with RAST (Overbeek et al., 2014). The online MAST server was used to identify tandem repeats (Bailey et al., 2009).

3.3.13 Phylogenetic analysis of orthologous proteins

54 bacterial genomes comprising a representative sample of insect symbionts and the *Enterobacteriaceae* genera *Tatumella*, *Erwinia*, *Pantoea*, and *Klebsiella* were downloaded from NCBI. Genomes were annotated with PROKKA version 1.5.2 as described previously. Annotated genomes were combined with the ErwSC and TA1 assemblies generated over the course of this study, along with the independent assembly of the *Ca.* E. dacicola metagenomic data from *B. oleae* midgut larval caeca ErwIsr. OrthoMCL was used to identify 11968 orthologous clusters (Li et al., 2003). Cluster occurrence was transformed in to presence-absence binary counts and heatmaps drawn from the Euclidian distance matrix of these counts in R

version 3.2.3 (R Core Team). A total of 81 single orthologs shared by all taxa were used to draw Maximum-Likelihood phylogenies. The amino acid sequences for each cluster were aligned with MUSCLE (Edgar, 2004), informative sites selected with Gblocks (Castresana, 2000), and the remaining alignments concatenated. Maximum-Likelihood trees were constructed with FastTree (Price et al., 2009). On the basis of this phylogeny, close relatives of ErwSC and TA1 were selected to repeat ortholog analysis and construct ML trees. This analysis comprised all *Erwinia* and *Tatumella* isolates from the previous analysis, BFo1 and BFo2 from Western Flower Thrips, *Erwinia chrysanthemi* (*Dickeya dadantii*) to root the tree, and ErwSC*,* ErwIsr and TA1. Phylogenetic analysis was repeated as above, this time using 826 single orthologous clusters shared by all taxa. In addition, OrthoMCL analysis was repeated with ErwSC, BFo1 and *E. amylovora*, in order to segregate phylogenetically related gene clusters involved in obligate symbiosis, facultative symbiosis, and free-living pathogenicity, respectively. The package VennDiagram was used to draw Euler diagrams of orthologous clusters in R version 3.2.3 (Chen and Boutros, 2011). Diagrams were annotated with the percentage of clusters unique to each organism as a proportion of the total numbers of clusters in each genome, as an indicator of the accessory genome associated with each lifestyle.

3.4 Results

3.4.1 Genome assemblies

Draft genome assemblies were generated for *Ca.* E. dacicola ErwSC, *Tatumella* TA1, *Chryseobacterium* CBo1, and *Stenotrophomonas* SBo1 (Figure 3.1). Assembly statistics for all individual *Ca.* E. dacicola assemblies indicate that a combination of single-cell and metagenomic datasets provides the highest quality assembly (Table 7.1). The quality-filtered ErwSC hybrid assembly comprised 333 contigs with 1000X coverage, had a total size of 2.1 Mb, and a GC content of 53.5 %. Most *Erwinias* have a GC content of 50-55 %, and a genome size ~4 Mb (Starr and Chatterjee, 1972). Genome reduction has occurred in *Ca.* E. dacicola, but the AT-bias common in the genomes of vertically transmitted bacteria is not present. The TA1 draft genome was

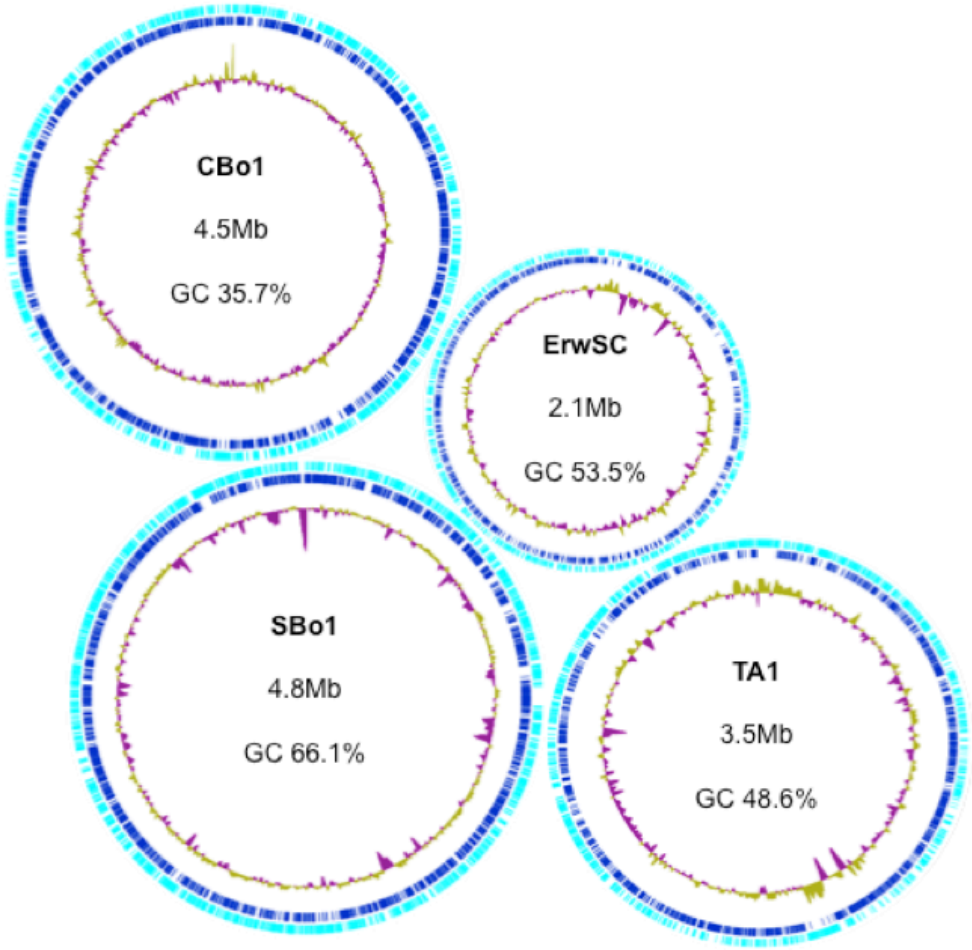


Figure 3.1 | Graphical representation of the draft genomes that were generated for taxa associated with wild *B. oleae*. Outer tracks show the forward CDS, dark blue tracks the reverse CDS, and inner tracks GC content. Assembly name, size and GC content are listed in the centre of each draft genome.

assembled into two contigs totalling 3.5 Mb in size with a GC content of 48.6 %, which is similar to other members of the *Tatumella* genus. All four draft genomes contained mobile genetic elements including bacteriophage, plasmids, and transposable elements. SBo1 and CBo1 have genomes 4-5 Mb in size, and expected GC content in comparison to close phylogenetic relatives. Examination of the draft genomes suggests that SBo1 and CBo1 represent allochtonous taxa that are able to survive within the insect gut, but do not necessarily have a functional interaction with the host.

3.4.2 Phylogenetic placement of *Ca.* E. dacicola and TA1

81 concatenated single orthologous proteins from 57 taxa including obligate and facultative insect symbionts, and putative close free-living phylogenetic relatives of *Ca. E. dacicola* and TA1 were used to estimate an ML phylogeny (Figure 3.2). *Ca. E. dacicola* clustered with the genus *Erwinia*, and TA1 with the genus *Tatumella*. This agrees with phylogenetic placement of these taxa by sequencing of the 16S rRNA gene. Closest relatives were the Western Flower Thrip (WFT) symbionts BFo1 and BFo2. Ortholog analysis was repeated with a smaller subset of closely related taxa, producing a set of 826 single orthologs. The resulting ML phylogeny had a fully supported topology that corroborates the placement of both taxa (Figure 3.3).

3.4.3 Adaptation to symbiosis

Euclidian distances of the presence and absence of orthologs grouped taxa by phenotype, with clear separation between obligate and facultative symbionts (Figure 3.4). *Ca.* E. dacicola clustered with the obligate insect endosymbionts, and TA1 with free-living and facultative bacteria. Taxonomic similarity defined the structure within these broad clusters. As in the phylogenetic reconstructions, TA1 remains closely affiliated with BFo2, whereas ErwSC and ErwIsr cluster away from BFo1. This reflects the difference in lifestyle between BFo1, which has retained the ability to be free-living, and *Ca*. E. dacicola, which is only found in association with *B. oleae*.

Figure 3.2 | Maximum-Likelihood phylogeny of 59 free-living and insect symbiotic members of the Gammaproteobacteria. Taxa were included in the analysis based on their phylogenetic proximity according to analysis of the 16S rRNA marker gene. The tree was inferred from a concatenated alignment of 81 orthologous clusters that were generated from an OrthoMCL analysis of complete or draft genomes and were present for all taxa included in the tree.

Ortholog analysis was conducted for ErwSC, BFo1, and *E. amylovora*, in order to ascertain the proportion of the core *Erwinia* genome retained by *Ca.* E. dacicola. Despite extensive genome reduction, 22.2 % of the ErwSC genome comprised accessory genes, similar to *E. amylovora* (20.8 %), and BFo1 (28.4 %) (Figure 3.5). RAST annotation of COG functional categories indicated that, in comparison to the other taxa isolated from the *B. oleae* gut, and taking into account genome size, the *Ca.* E. dacicola genome is enriched in features involved in amino acid metabolism and biosynthesis, and deficient in those related to carbohydrate metabolism (Figure 3.6).

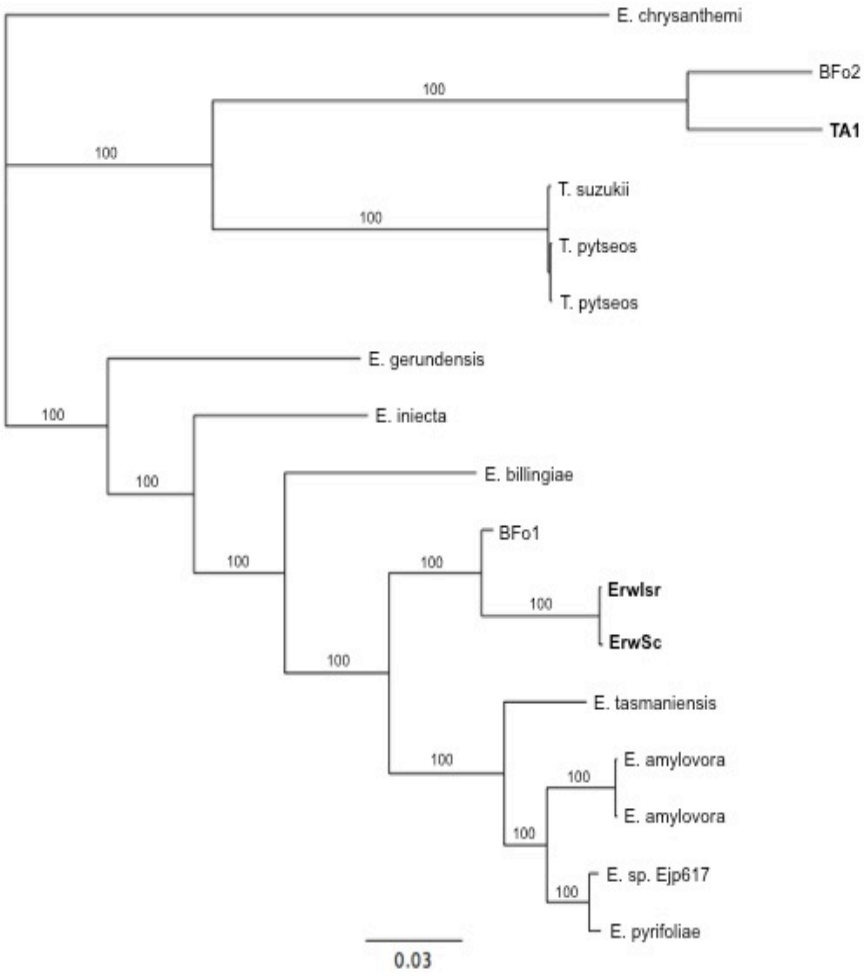


Figure 3.3 | Maximum-Likelihood phylogeny of *Erwinia* and *Tatumella* strains, including *Ca.* E . dacicola and TA1 from *B. oleae*. The tree was drawn from a concatenated alignment of 826 orthologous clusters that were generated by OrthoMCL analysis of complete or draft genomes for all taxa included in the tree.

A small but significant portion of the accessory genome is dedicated to nitrogen metabolism. *Ca.* E. dacicola has retained several operons that may be involved in adaptation to conditions experienced within the gut of *B. oleae*. *Rhodobacter* nitrogen fixation (rnf) operons are often found in strict or facultative anaerobes where they enable the utilisation of ferredoxins as electron donors (Biegel et al., 2010). However, in some members of the *Enterobacteriaceae* including *Buchnera*, the rnf operon is involved in the response to oxidative stress (Koo et al., 2003; Charles et al., 2011). The production of Reactive Oxygen Species (ROS) upon activation of the immune system can induce severe oxidative stress in the insect gut (Ha et al., 2005).

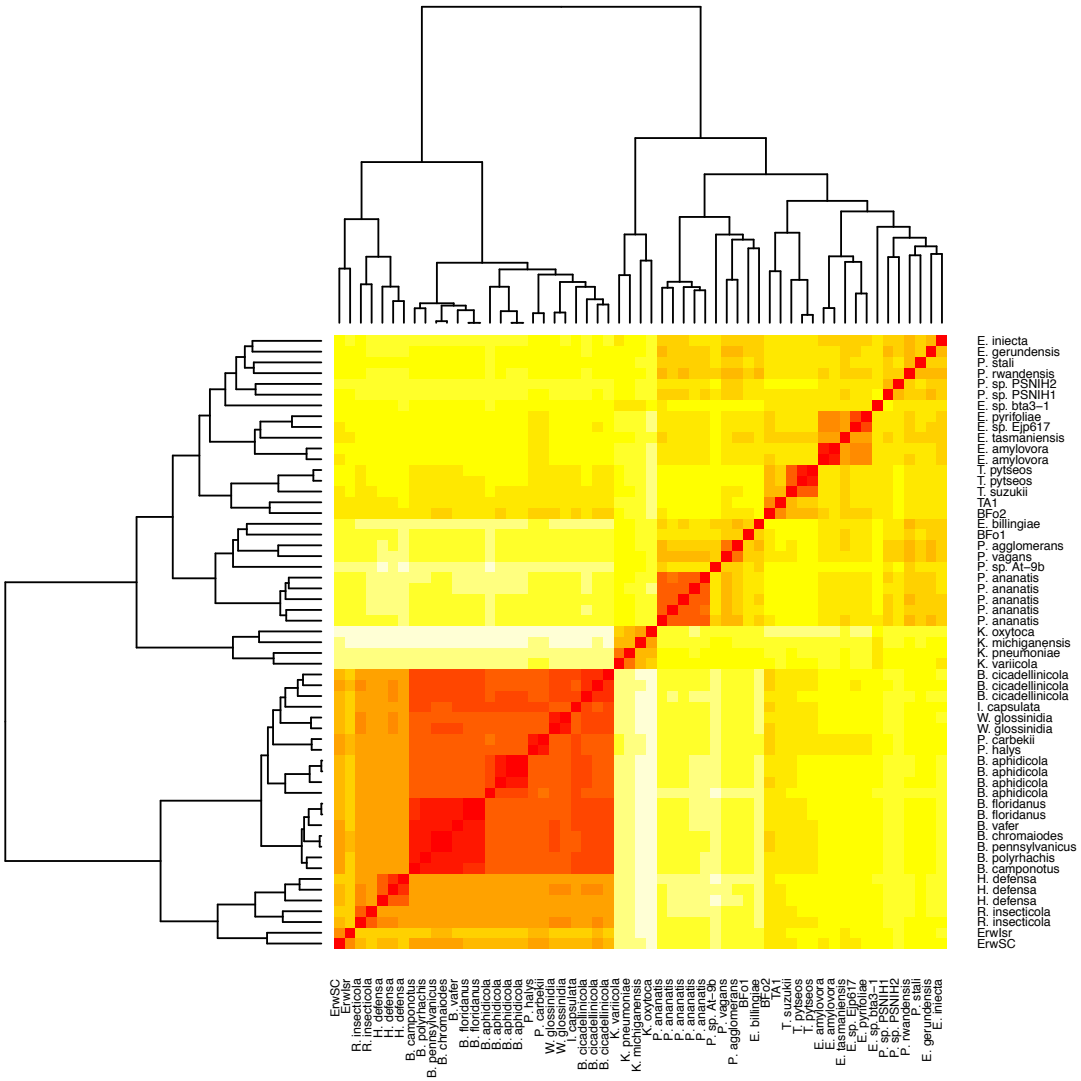


Figure 3.4 | A heatmap illustrating the hierarchical clustering of 11968 orthologous clusters present in 59 free-living and insect symbiotic members of the Gammaproteobacteria. Clustering is based on the Euclidian distance matrix of the presence and absence of orthologous clusters in all 59 taxa, including *Ca.* E. dacicola and TA1 from *B. oleae*. Darker colours indicate shorter distances and lighter colours longer distances between taxa. Functional capacity is not strictly congruent with the phylogeny inferred from a smaller number of shared orthologs in Figure 3.2.

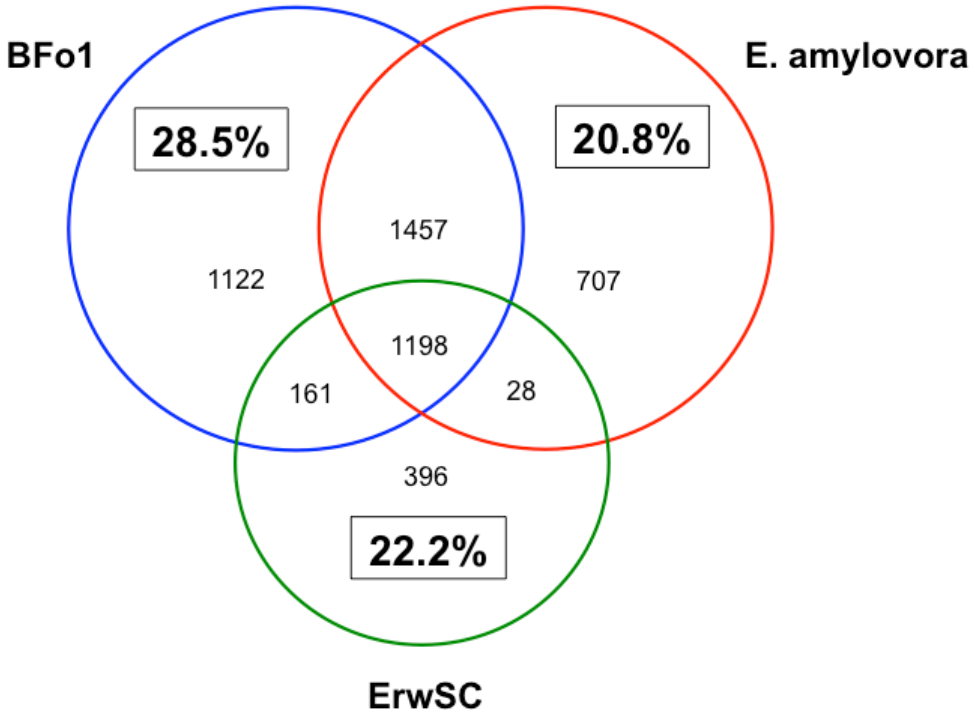


Figure 3.5 | Euler diagram displaying the number of orthologs shared by the free-living *E. amylovora*, facultative symbiont of thrips BFo1, and obligate symbiont of *B. oleae Ca.* E. dacicola. Percentages represent the proportion of clusters unique to each genome and are analogous to the accessory genome. Despite being close phylogenetic relatives, each bacterial genome houses 20-30% unique genes, indicating that these genes may be associated with a specific lifestyle.

Ca. E. dacicola and TA1 have complete rnf operons that are phylogenetically distinct (65 % amino acid similarity rnfC), indicating that they have not recently been co-opted by horizontal transmission. CBo1 and SBo1 possessed none of the rnf genes, despite SBo1 being in the same family as *Ca.* E. dacicola and TA1. The Iron Sulfur Cluster (isc) and Sulfur Mobilization (suf) operons are well characterised in the pathogenic *E. chrysanthemi*, where they mitigate oxidative stress and iron limitation of host origin, respectively. Both operons encode Fe-S clusters, which are highly versatile cofactors of proteins. Fe-S clusters serve a number of functions within the cell including DNA repair, metabolism, electron transport, and RNA modification (Rincon-Enriquez et al., 2008). CBo1 has neither the suf or isc operon, SBo1 has only the suf operon, and both *Ca.* E. dacicola and TA1 have both. *Ca.* E. dacicola and TA1 also

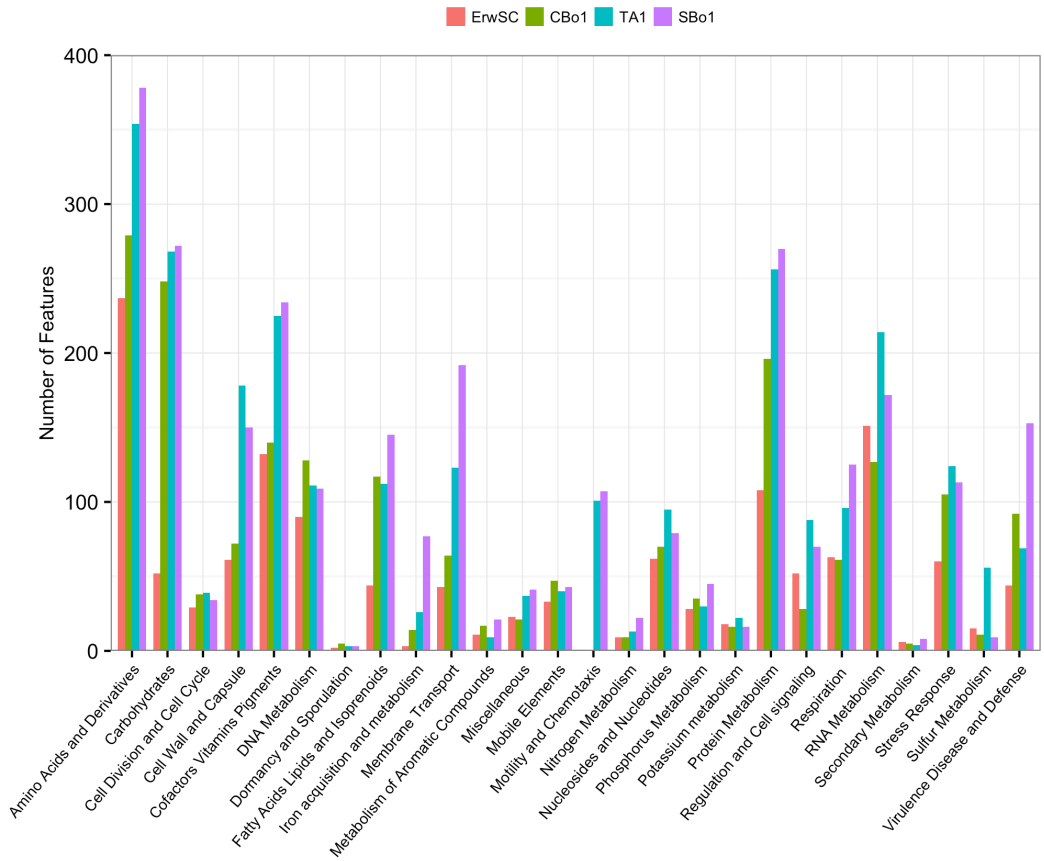


Figure 3.6 | The number of features assigned to COG functional categories by RAST in the four draft genomes that were generated for taxa isolated from wild *B. oleae*. ErwSC = *Ca*. E. dacicola; CBo1 = *Chryseobacterium*; TA1 = *Tatumella* TA1; SBo1 = *Stenotrophomonas*.

encode the *suhB* gene as part of the isc operon, which has been shown to reduce the cold sensitivity of cells when expressed (Wang et al., 2007).

3.4.4 Nitrogen recycling

Metabolic pathway reconstruction from genome sequences indicates that nitrogen-recycling pathways previously characterised in other insects and their endosymbionts may be employed by *B. oleae* and its microbiota (Figure 3.7). Excluding allantoicase (EC 3.5.3.4), all of the required enzymes for the generation of glutamine from uric acid were detected in either *Ca.* E. dacicola or TA1. Critically, neither microbe has retained all steps, and therefore could not employ this pathway independently. The first step in the assimilation of host waste uric acid to ammonia, which can be incorporated back in to metabolism, is the conversion of uric acid to allantoin via the enzyme uricase.

Uricase enzymes were detected in all taxa studied, excluding *Ca.* E. dacicola, indicating that this gene has been lost either through drift or purifying selection, and that it is therefore not essential for the association.

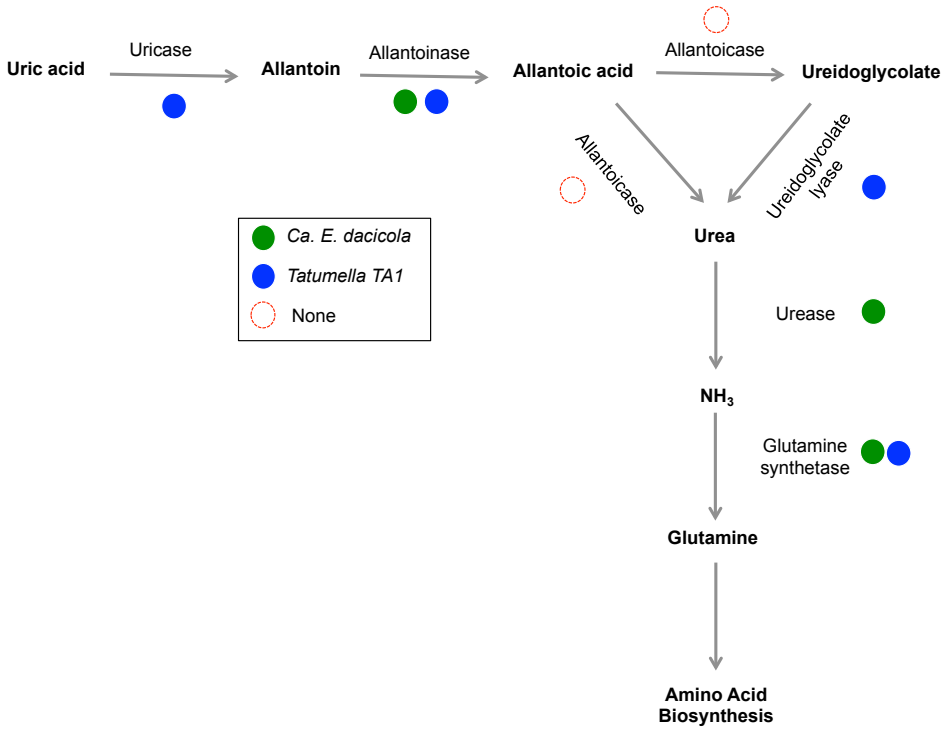


Figure 3.7 | The distribution of genes associated with the utilization of host waste nitrogenous products in the genomes of *Ca.* E. dacicola and TA1. The presence and absence patterns of annotated genes suggest that nitrogen-recycling pathways may be shared between *Ca.* E. dacicola and TA1.

Another indication that TA1 forms a permanent association with *B.* oleae, and that it may participate in metabolic crossfeeding with *Ca.* E. dacicola, is that it encodes the genes to convert ureidoglycolate to urea. In addition, the *glpF* gene that encodes a urea membrane transport protein is located <100 kb downstream. Generation and export of urea is a key step in the pathway. *Ca.* E. dacicola encodes a urease operon, which is characteristic of endosymbionts involved in nitrogen recycling. Ureases convert urea in to carbon dioxide and ammonia. The urease operon comprises three structural genes *ureABC*, five accessory genes *ureDEFGJ*, and a transcriptional regulator *ureR* (Mulrooney and Hausinger, 1990; D'Orazio and Collins, 1993). The ErwSC assembly encodes *ureABCDEJ*, whereas ErwIsr encodes *ureABCDEFGJ*, it has not been ascertained whether this is due to an error in

the assembly and annotation process, or whether they are true strain-wise differences. No other *Erwinia* genomes sequenced to date house urease operons. *ureR* was not annotated in the ErwIsr or ErwSC draft genomes, but was present in BFo1 and BFo2, and in a truncated form in TA1. Determining whether the host or another member of the microbiota such as a fungal symbiont encodes allantoicase is a critical step in deciphering the metabolic cross feeding that may mediate nitrogen recycling in *B. oleae*.

3.4.5 TA1 adhesion proteins

One advantage of using long-read sequencing technologies such as PacBio is to bridge highly repetitive regions and elucidate their size and structure. The TA1 genome encodes a giant IgA-rich outer-membrane protein similar to SiiE proteins utilised by *Salmonella enterica* to promote adhesion to host cells (Griessl et al., 2013). The protein is 10,288 amino acids in length, and contains 59 copies of a 90 bp Ig-like Big34 domain repeat that shows little divergence in amino acid sequence between repeats (Figure 3.8). These characteristics are shared with other large surface proteins, along with threonine rich composition and a lack of cysteine residues (Reva and Tümmler, 2008).

Figure 3.8 | The conserved repeat identified by MAST within the Big34 domains of a putative 30kb adhesion protein in TA1, which is associated with a larger putative adhesion operon containing a Type I secretion system. Position numbers are relative to the start of the consensus sequence of the repeat motif detected by MAST, and the size of the letters indicates the degree of conservation within the repeat.

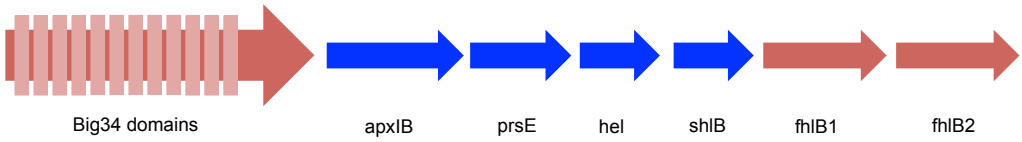


Figure 3.9 | A putative adhesion operon identified in the closely phylogenetically-related TA1 from *B. oleae* and BFo2 from thrips comprising secreted surface proteins that promote adhesion in *S. enterica* and *E. chryseobacterium*, and containing the giant adhesion-like proteins that contain repeating Big34 domains. Big34 = putative Big34 repeat-containing adhesion protein; apxIB = putative ABC transporter/TISS; prsE = TISS; hel = outer membrane protein; shlB = hemolysin transporter; fhlB = filamentous hemagglutinin. Elements coloured in red are putative adhesion elements of the operon, and elements in blue are putative secretion elements of the operon.

The Big34 region is flanked by *apxIB*: an ABC transporter of bacteriocins against closely related bacteria that interacts with membrane proteins to form type I secretion systems (Lin et al., 2015), and *prsE*, which encodes a type I secretion system protein. *Salmonella* SiiE is secreted to the cell surface by a Type I secretion system (TISS) encoded on pathogenicity island 4 in *S. enterica* (Morgan et al., 2007). In addition, in TA1, genes encoding an outer membrane protein (*hel*), hemolysin transporter (*shlB*) and filamentous hemagglutinin (*fhlB*) are immediately downstream of the Type I secretion system proteins (Figure 3.9). All of the components of this operon were detected in BFo2. However, the BFo2 draft genome was generated with Illumina short-read sequencing technology, and is therefore highly fragmented in comparison to the TA1 draft genome. Therefore, synteny is not observed between the two assemblies. In addition, the Big34 proteins were disparate in size, number of repeats, and the homogeneity of repeats (Figure 3.). This is most likely due to the inability of short-read sequencing technologies and their associated assemblers to reconstruct large numbers of short tandem repeats accurately.

Figure 3.10 | The conserved repeat identified by MAST within the Big34 domains of a putative adhesion protein in BFo2, which is associated with a larger putative adhesion operon that is also found in TA1. Position numbers are relative to the start of the consensus sequence of the repeat motif detected by MAST, and the size of the letters indicates the degree of conservation within the repeat.

3.5 Discussion

The *Ca.* E. dacicola draft genome is reduced in size in comparison to free-living Erwinias, but maintains a comparable GC content (Starr and Chatterjee, 1972). *Ca.* E. dacicola is not restricted to one organ throughout the life cycle, is vertically transmitted via egg smearing on the surface of embryos rather than transovarially, and can exist both intra and extracellularly (Estes et al., 2009). For these reasons, *Ca.* E. dacicola is likely to achieve larger population sizes than endosymbiotic bacteria, which suffer from Muller's ratchet as a result of reduced effective population size (Moran, 1996). Free movement in the gut likely generates greater admixture, possibly with other members of the gut microbiota, and enables purifying selection to preserve optimised codon usage. Due to the inherent deletion bias in bacterial genomes, genome reduction still occurs in this host-associated bacterium (Mira et al., 2001). This acts to remove redundant proteins that provide no fitness benefit in the context of the symbiosis, and streamlines the genome to host-associated functions (Hansen and Moran, 2011). Genome reduction and co-cladogenesis with the host are common features amongst other gut-associated vertically transmitted symbionts of insects, even though they are not intracellular (Hosokawa et al., 2006).

Previous studies have implicated the role of *Ca.* E. dacicola in amino acid biosynthesis in larvae and nitrogen provisioning in adults (Ben-Yosef et al., 2014; 2015). Lysine is an essential amino acid that is depleted in ripening olives due to protein cross-linking by phenolic compounds. Larval development in ripening olives is dependent upon the presence of the native microbiota, which is hypothesised to synthesise amino acids (Ben-Yosef et al., 2015). Both the *Ca.* E. dacicola and TA1 genomes encode the entire lysine biosynthesis pathway, indicating that either, or both, may provision the host with lysine.

The native microbiota also plays a nutritional role in adults, which feed on a polyphagous diet of bird faeces, bacteria, honeydew, and plant exudates

(Bateman, 1972; Fletcher, 1987; Drew and Yuval, 2000). As in larvae, the fitness contribution of the microbiota is diet-dependent, and is only observed in nutrient-poor conditions when essential amino acids are limiting (Ben-Yosef et al., 2010; 2014). Examination of the metabolic pathways present in *Ca.* E. dacicola and TA1 indicate that they encode proteins that may be able to assimilate the host nitrogenous waste product uric acid to glutamine, which can be repurposed back in to amino acids. However, there were no detectable allantoicase enzymes encoded either by the host or any member of the microbiota studied. One possible explanation is that allantoicase is encoded by another member of the microbiota that has not yet been isolated. It is possible that fungal or yeast-like symbionts may provide these enzymes, as has been demonstrated in *Nilaparvata lugens* (Hongoh and Ishikawa, 1997). Or, it may be that the enzyme was simply not identified due to the methodology employed, either because the enzyme was not included in the existing genome assemblies, was not identified due to a lack of homology in the databases employed, or because the organism that encodes the enzyme was not studied.

Allantoicase is a key step in the proposed pathway, as it facilitates the conversion of allantoic acid to urea and ureidoglycolate. In cockroaches and ants, *Blattabacterium* and *Blochmannia* do not encode allantoicase. Instead, both insect genomes encode allantoicase enzymes (Patino-Navarrete et al., 2014). The nature of these symbioses likely facilitates this mode of action, as the symbionts are housed in bacteriocyte organs where hosts can tightly control expression and nutrient exchange. In other insect-bacterial symbioses that depend upon metabolic cross feeding, genes that facilitate insect-bacterial interactions are up regulated in bacteriocytes (Hansen and Moran, 2011). However, the gut represents a less stable environment with higher microbial diversity and greater potential for metabolic redundancy. Allantoicase enzymes are more common amongst fungi than insects or bacteria. Fungal diversity has not previously been assessed in the *B. oleae* gut, but fungi are commonly associated with insects (Gibson and Hunter, 2010). Therefore, further investigation in to the presence of an allantoicase enzyme, possibly of fungal origin, is warranted.

B. oleae experiences seasonal variation in environmental conditions, and in the availability and quality of dietary nitrogen. The ability of the microbiota to assimilate ammonia from stored forms of nitrogen such as uric acid and urea provides an evolutionary strategy to mitigate these fluctuations. *B. oleae* is multivoltine, producing up to five generations per year in some locations, and not all generations experience the same sources or availability of dietary nitrogen (Tzanakakis, 2003). *Olea europaea*, the main host of *B. oleae*, flowers in late spring and produces fruits from early summer in to the winter, which can remain on trees until the following spring. Larvae require the native microbiota for development in ripening, but not ripened olives (Hagen, 1966; Tzanakakis and Stavrinides, 1973; Lambrou and Tzanakakis, 1978; Ben-Yosef et al., 2015). Similarly, when essential amino acids are provided in the diet, adults are not dependent upon their native microbiota for reproduction (Ben-Yosef et al., 2010; 2014). Diet-dependent responses may reflect variation in natural sources of nitrogen such as pollen, which are rich in vitamins and amino acids but are only available at certain times of the year (Roulston and Cane, 2000).

Nitrogen is a critical resource for reproductive development in insects (Awmack and Leather, 2002). In *B. oleae*, adults continue to reproduce over the winter despite low temperatures (Tzanakakis, 2003), but undergo reproductive diapause during the summer when oviposition substrates are lacking and environmental conditions such as high temperatures and low humidity are not permissive to reproduction (Fletcher et al., 1978; Economopoulos et al., 1982). Symbiotic bacteria were previously hypothesised to manage the exit of females from reproductive diapause in response to fruit availability, though a mechanism has not been proposed (Koveos and Tzanakakis, 1990; 1993). The findings of this study suggest that the microbiota may be able to prime the nutritional status of the fly by mobilizing uric acid reserves prior to the emergence of new fruit. Symbionts able to use these sources of nitrogen (*Ca.* E. dacicola and TA1) would have a replicative advantage, increasing their relative abundance. An increase in symbiont density may promote vertical transmission to the next generation by

egg smearing (Jaenike, 2009), ensuring early infection of larvae and the essential presence of the symbiont in juveniles developing in ripening olives.

TA1 encodes proteins orthologous to those that play a key role in host cell attachment and biofilm formation in *Salmonella* (Latasa et al., 2005). Orthologous proteins were found in the genomes of other members of the genus *Tatumella*, indicating that they are ancestral features. Intracellular bacteria inhabit the cells of *B. oleae* larval midgut caeca, and biofilms form in the oesophageal bulbs of adults within 48h of emergence (Estes, 2009). Biofilms enable more efficient metabolic cross feeding, and act to stabilise interactions in environments that are permissive to nutrient exchange (Hoek et al., 2016). For example, deletion of genes that are essential for biofilm formation in environmentally acquired *Burkholderia* gut symbionts of *Riptortus* beanbugs results in disintegration of the symbiosis and a subsequent reduction in host fitness (Kim et al., 2014). Two morphologies of Gram-negative enteric bacteria have been identified in these tissues, both of which are presumed to be *Ca.* E. dacicola (Estes et al., 2009). However, the culture-independent techniques that were used to identify bacteria in this study (FISH, 16S PCR) can be insensitive to low-abundance taxa, and as TA1 is generally found at less than 5 % relative abundance it may have been overlooked by these approaches.

TA1 has been isolated from adults that emerged in a sterile laboratory environment, indicating that it is capable of surviving metamorphosis, as is *Ca.* E. dacicola. It is not yet clear how either bacterium achieves this, but surface adhesion proteins that are able to attach to host cells such as midgut progenitor cells in larvae may provide part of the explanation. They may also play a role in vertical transmission from one generation to the next. Both *Ca.* E. dacicola and TA1 encode features that enable them to adapt to physiological conditions encountered within a host environment or host cells. These include the Suf and Isc Fe-S cluster biogenesis operons, which allow bacteria to adapt to oxidative stress and iron limitation, respectively (Rincon-Enriquez et al., 2008). These operons are similarly retained in the severely reduced genomes of several symbionts of sap-feeding insects, indicating that

they may be a convergent feature amongst insect symbionts (McCutcheon and Moran, 2012; Sabree et al., 2013).

The close phylogenetic relationship between the Western Flower Thrip (WFT) symbionts BFo1 and BFo2, and the *B. oleae* symbionts *Ca.* E. dacicola and TA1 is particularly striking. The disparity in genome size between *Ca.* E. dacicola and BFo1 suggests that *Ca.* E. dacicola has been host-associated for a longer period of time (Facey et al., 2015). BFo2 and *Tatumella TA1* represent a novel branch of the *Tatumella* genus in comparison to the draft genomes that are currently available. Both are found at frequencies comparable to those observed in insect facultative symbionts (Chanbusarakum and Ullman, 2008). In addition to the phylogenetic origins of their symbiotic bacteria, *B. oleae* and the WFT share several ecological characteristics. Both feed on plant-based diets and switch from monophagy to polyphagy as they develop from juveniles in to adults, which feed on pollen in addition to the plant leaves used by larvae (Mound and Teulon, 1995). Symbiont contribution to fitness is diet-dependent, and correlates with fluctuations in nitrogen availability in both organisms (de Vries et al., 2004; Ben-Yosef et al., 2010). Though it is known that the microbiota of WFT contributes to development and fecundity in a diet-dependent manner, the individual roles of BFo1 and BFo2 have not yet been determined (de Vries et al., 2004). Further genomic comparisons of *Ca.* E. dacicola and BFo1, and TA1 and BFo2 may elucidate their functional and evolutionary relationships, and the roles that they play within their respective hosts.

In conclusion, genomic analyses indicate that *Ca.* E. dacicola and TA1 are host-adapted bacterial symbionts that are able to provision the host with essential amino acids and nitrogen. These findings agree with experimental data from previous studies, and are analogous to nitrogen recycling mechanisms in other insect-bacterial associations. The phylogenetic and phenotypic similarities between *B. oleae* and the WFT symbionts warrant further investigation, and indicate that this pairing may have evolutionary significance beyond the *B. oleae* gut. The discovery of rare giant surface adhesion proteins and an accompanying toxin secretion operon in TA1

suggests that it may be able to invade host cells. In addition, surface adhesion proteins may facilitate biofilm formation and more efficient nutrient exchange, which could act to stabilise the association between all three members. The discovery of the potentially novel symbiont TA1 may also provide an explanation for the non-ubiquitous nature of *Ca.* E. dacicola in natural populations (Estes et al., 2012a). Further work is required to determine how metabolic processes are partitioned between *B. oleae* and its associated microbes, and therefore how different members of the microbiota contribute to host fitness.

Chapter 4 : Effect of Bacterial Infection on *Bactrocera oleae* Fitness and Behaviour

I designed and conducted the experiment, performed the data analysis, and wrote the chapter.

4.1 Abstract

Background: There is strong evidence that the composition, structure, and transmission of the microbiota of *Bactrocera oleae* are drastically altered by artificial rearing practices. Changes in diet, the use of preservatives, and a shift in the environmental microbes that are present in the laboratory as opposed to the natural environment are all thought to contribute to this change. However, it remains unclear whether these processes have any effect on *B. oleae* health.

Methods: 16S rRNA amplicon profiling was used to assess how community composition changes over the course of the *B. oleae* life cycle under mass-rearing conditions. Perturbation of the microbiota was also assessed experimentally in adults. The microbiota of mass-reared insects was altered by treatment with antibiotics, and by supplementation of the diet with bacteria isolated from wild-caught *B. oleae*. Wild flies were subjected to antibiotic treatment to assess the impact of removal of the native microbiota on host health. Fly health was measured by activity monitoring, which allowed us to gather continuous measurements of activity and longevity over a period of several days after alteration of the microbiota.

Results: Both community profiling and manipulation of the microbiota indicated that the *B. oleae* mass-reared microbiota is susceptible to perturbation. Community composition and structure varied significantly between juveniles and adults in a pattern that is not observed in wild flies. Experimental manipulation of bacterial communities elicited strain-specific and sexually dimorphic responses in host behaviour.

Conclusions: Removal of the native microbiota through mass-rearing practices may make the *B. oleae* gut more susceptible to invasion by microbes that are able to impact upon fly health. Future strategies to improve mass-rearing strategies should be designed to stabilise the transmission of commensal bacteria between life stages, which may protect against opportunistic pathogens and improve rearing efficiency.

4.2 Introduction

Many insect species are holometabolous, undergoing significant morphological changes over the course of their life cycle (Truman and Riddiford, 1999). Decoupled morphological adaptations between life stages allow insects to exploit a broader range of habitats and food sources. They also facilitate ontogenetic specialisation by life stage-specific selection (Moran, 1994). However, metamorphosis presents a significant challenge to the microbes associated with insects. During metamorphosis, the gut – which is where the majority of bacteria that contribute to tephritid fitness reside (Lauzon, 2003), is excised and remodelled. In order to maintain mutualistic associations, the microbiota must either retain its association with the host throughout metamorphosis, or be reacquired from the environment upon ecdysis. Previous work by Estes *et al.* posits that *Ca.* E. dacicola is able to colonize larval midgut progenitor cells during metamorphosis and re-colonize the newly formed adult gut (Estes et al., 2009). This mechanism benefits both host and microbe by safeguarding the metabolic association for which they are both adapted; providing a habitat for the obligate *Ca.* E. dacicola (Capuzzo et al., 2005; Estes et al., 2009); and populating the vulnerable niche of the *B. oleae* gut with commensal microbes from the onset of adulthood.

The absence of an established microbiota, as is the case in newly-ecdysed adults, renders the gut highly susceptible to infection by opportunistic microbes (Greenberg and Klowden, 1972). Therefore, it is in the interest of both host and bacterium to ensure that the niche is robust (or tolerant) to invasion. The lepidopteran *Galleria mellonella* and *Enterococcus mundtii* have evolved mechanisms to protect this niche over the course of metamorphosis (Johnston and Rolff, 2015). Expression of immune enzymes by the host and direct antagonism of infecting microbes by *E. mundtii* ensure symbiont transmission while simultaneously excluding pathogenic microbes. An analogous strategy may be adopted by *B. oleae* and its gut microbiota to mitigate costly opportunistic infections.

Enabling vertical transmission of beneficial microbes is one strategy that holometabolous insects use to avoid colonisation by opportunistic pathogens. While some insects such as *B. oleae* are associated with the same microbes throughout their life cycle, not all insects form tight associations with bacteria during all stages of development. Several studies have identified trans-stadial variation in the structure and composition of insect-associated bacterial communities, where differences between life stages often reflect variation in diet and physiological conditions (Hammer et al., 2014; Andongma et al., 2015; Chen et al., 2016; Staudacher et al., 2016). Whilst multiple inoculations of the gut throughout development allow flexibility in response to environmental conditions, they also increase the probability of pathogen invasion due to frequent disruption. This is particularly pertinent in laboratory-reared insects, where artificial rearing practices can remove naturally adapted communities (Estes et al., 2012a; Wong et al., 2013; Morrow et al., 2015; Staudacher et al., 2016). Removal of these communities exposes niches and creates opportunities for inoculation that would otherwise be exploited by the native microbiota (Salem et al., 2015a). For example, the utilisation of rich and homogeneous artificial diets may remove the nutritional selection pressures that bacterial communities are exposed to in wild populations, resulting in the alteration of their gut microbiota. This leaves them susceptible to colonisation by maladaptive or pathogenic microbes.

There has been a recent revival of interest in the development of the Sterile Insect Technique (SIT) for *B. oleae* (Estes et al., 2012b). Significant efforts were made to develop an SIT program for *B. oleae* in the 1960s, 70s and 80, but were abandoned due to the difficulty of rearing large numbers of *B. oleae* (Economopoulos et al., 1982). Artificial diets in particular were highlighted as an area that required improvement in order to increase the rearing success and fitness of adults for release in SIT programs (Estes et al., 2012b). Since then, improved methods for the sterilisation of insects have been invented, including the transgene-based method Female-Specific Release of Insects with a Dominant Lethal (fsRIDL). An fsRIDL strain of *B. oleae* demonstrated significantly improved mating success compared with irradiated *B. oleae* strains, which was attributed to synchronous mating with wild insects, and

induction of mating refractoriness in wild females (Ant et al., 2012). However, problems with rearing still apply to fsRIDL *B. oleae*, particularly as fsRIDL depends upon the use of antibiotics for suppression of the female-lethal transgene during rearing to build up large populations of insects for release (Thomas, 2000)

Anecdotal evidence of severe population crashes is common in *B. oleae* rearing (Oxitec Ltd.- Unpublished). It is thought that pathogenic microorganisms may cause these population crashes, which invade due to mass-rearing practices that leave the gut exposed through constant disruption. The discovery of the culture-viable, putative mutualist *Tatumella* TA1 provides the opportunity to supplement the *B. oleae* laboratory microbiota with a mutualistic bacterium that can be transformed to be resistant to antibiotics (Chapter 3). TA1 was isolated from the guts of wild *B. oleae*, and has previously been found at low relative abundance in *B. oleae* populations from Crete and Israel (Chapter 2, Chapter 3). Whole genome sequencing indicates that TA1 may be able to supply the host with essential amino acids, and may also be able to invade host cells (Chapter 3). We therefore decided to test whether the inoculation of adult or juvenile laboratory-reared *B. oleae* would result in a consistent infection with TA1, and whether this had any effect on *B. oleae* health. Kounatidis *et al.* successfully infected laboratory-reared insects with *Acetobacter tropicalis*, which formed a biofilm-like structure in association with the peritrophic matrix (Kounatidis et al., 2009). When primed with a *B. oleae*-adapted microbe, the gut may be more robust to perturbation, which would benefit host health by reducing the risk of infection and avoiding the activation of costly immune responses.

Locomotor activity is commonly used as an indicator of health in both invertebrates and vertebrates. Increased somnolence and lethargy in sick individuals is a behavioural adaptation to infection (Hart, 1988). By reducing their activity, sick individuals may compensate for the metabolic costs of infection (Ardia et al., 2012), or increase predator avoidance by sheltering and avoiding detection when debilitated (Dantzer, 2004; Dantzer and Kelley, 2007; Otti et al., 2011; Ardia et al., 2012). Similarly to pathogenic or parasitic

organisms, mutualists must obtain their resources from the host environment, and therefore can be costly to maintain. Therefore, they may only be of net benefit to host fitness under certain environmental conditions. In *Drosophila melanogaster*, *Wolbachia* titre is directly related to virulence. The over-proliferating *Wolbachia* strain *w*MelPop increases mortality as density increases (Chrostek and Teixeira, 2015). Even the naturally protective strain *w*Mel, which defends *Drosophila* hosts against Drosophila C Virus and other RNA viruses (Hedges et al., 2008; Teixeira et al., 2008), reduces host longevity (Chrostek et al., 2013; 2014). In addition to reducing life span, *Wolbachia* infection can make males less active (Vale and Jardine, 2015).

Both activity and longevity are critical factors in the success of SIT programmes, as they depend upon the ability of males to survive and compete for matings. Therefore, both are important parameters to measure when developing rearing programs for sexually competitive SIT strains. We used culture-based techniques to assess the effect of antibiotic treatment and inoculation of a GFP-transformed TA1 strain on bacterial communities, and activity monitoring to assess the corresponding host response. In addition, 16S rRNA gene amplicon profiling of the bacterial communities associated with each life stage of artificially reared *B. oleae* revealed a large shift in community composition between juveniles and adults, which corresponds with a significant change in diet composition. By assessing the role of bacteria in the health of laboratory-reared *B. oleae*, we hope to contribute to the improvement of mass-rearing programmes for SIT.

4.3 Methods

4.3.1 Fly stocks

Argov strain flies (Ant et al., 2012) were employed for all 16S rRNA amplicon profiling and activity monitoring experiments. Wild flies that were employed for activity monitoring experiments were sampled from olive trees in the grounds of the University of Crete (Heraklion, Greece) in November 2015 by collecting infested olives and allowing third instar larvae to emerge and pupate in sterile sand. Adult flies were reared following conventional laboratory practices (Genc and Nation, 2008; Appendix 1).

4.3.2 16S rRNA gene amplicon profiling

Individuals were collected at all stages of the life cycle and stored immediately at -80 ^{o}C until DNA extraction. Insects were processed individually, excluding eggs, which were collected in pools of approximately 300 and resuspended in 30 µl molecular water. Teneral adults were sampled within 24 h of emergence, and sexually mature adults at 7-days post emergence. DNA was extracted from individual insects using the Qiagen DNeasy Blood and Tissue protocol for Gram-positive bacteria (Qiagen, Manchester, UK) following the manufacturers' instructions. An additional bead-beating step with 3 mm carbide beads (Qiagen, Manchester, UK) at 25 Hz for 30 seconds in a Qiagen tissue lyzer (Qiagen, Manchester, UK) was employed prior to digestion with Proteinase K.

To approximate the relative abundance of *Ca.* E. dacicola in wild larvae, the restriction fragment polymorphism assay developed by Estes *et al.* for *Ca.* E. dacicola 16S rRNA gene PCR amplicons was employed (Estes et al., 2014). A 1500 bp region of the 16S rRNA gene was amplified with universal primers 8F (5'–AGAGTTTGATCMTGGCTCAG–3') and 1492R (5'-CCCCTACGGTTACCTTGTTACGAC–3'). Reactions were performed in a total volume of 25 µl comprising 12.5 µl MyTaq Red (Bioline), 0.5 µl of each 10 µM primer stock, 1 µl template gDNA, and 10.5 µl PCR-clean water. Thermal cycling conditions were 95 ^{o}C for 5 minutes, 30 cycles of 95 ^{o}C for 30

seconds, 56 °C for 45 seconds, 72 °C for 90 seconds, and a final extension at 72 °C for 7 minutes. Digestion with PstI was performed by incubating 500 ng 16S rRNA PCR product with 0.5 µl PstI-HF enzyme (New England Biolabs), 2 µl 10X Cut Smart buffer (New England Biolabs), and PCR-clean water in a final volume of 20 µl. The positive controls employed for the digestion step were *Arsenophonus*, which does not contain the restriction motif, and a pool of *B. oleae* adults where 85 % of the community is composed of *Ca.* E. dacicola, as verified by 16S rRNA gene amplicon sequencing (Chapter 3).

Due to the low concentration of bacterial DNA in juvenile insects and newly emerged adults, an initial round of PCR was performed using universal 16S primers UFPL (5'-AGTTTGATCCTGGCTCAG-3') and URPL (5'-GGTTACCTTGTTACGACTT-3'). PCR reactions were carried out in a volume of 25 µl containing 1 µl of template DNA, 12.5 µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs), 0.5 µl of each primer at 10 mM concentration and 10.5 µl PCR-clean water. Thermo-cycling conditions were 98 °C for 60 seconds, 25 cycles of 98 °C for 60 seconds, 56 °C for 45 seconds, 72 °C for 60 seconds, and a final extension at 72 °C for 10 minutes. 1 µl of this PCR product was then used as the DNA template in a second round of PCR for the amplification of the V4 region with fusion primers F515 (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso et al., 2011). The 806R reverse primers contained a unique Golay barcode specific to each sample for read de-multiplexing. PCR reactions were carried out in a volume of 25 µl containing 1 µl of template DNA, 12.5 µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs), 1 µl of each primer at 3 mM concentration and 9.5 µl PCR-clean water. Thermal cycling conditions were 30 seconds at 98 °C, 25 cycles of 98 °C for 10 seconds, 58 °C for 30 seconds and 72 °C for 30 seconds, followed by a final extension of 5 minutes at 72 °C. All PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter Genomics) and quantified with the Qubit dsDNA High-Sensitivity assay (Life Technologies), and an Agilent Bioanalyzer High-Sensitivity DNA chip (Agilent). Samples were pooled at equimolar concentrations and size-selected in a range of 350-450 bp by Pippin-Prep (Sage Science) if required.

Sequencing was performed at the University of Liverpool Centre for Genomic Research on an Illumina MiSeq platform with V2 chemistry, generating paired-end 250 bp reads (Illumina, San Diego, CA).

Raw sequencing reads were de-multiplexed and converted to FASTQ format using CASAVA version 1.8 (Illumina 2011). Cutadapt version 1.2.1 (Martin, 2011) was used to trim Illumina adapter sequences from FASTQ files. Reads were trimmed if 3 bp or more of the 3' end of a read matched the adapter sequence. Sickle version 1.200 (Joshi and Fass, 2011) was used to trim reads based on quality: any reads with a window quality score of less than 20, or were less than 10 bp long after trimming, were discarded. BayesHammer was used to correct reads based on quality (Nikolenko et al., 2013). Paired-end reads were assembled with a minimum overlap of 50 bp and discarded if outside of the range 200 to 300 bp using PandaSeq (Masella et al., 2012). All subsequent analyses were conducted in QIIME version 1.8.0 (Caporaso et al., 2010b). Sequences were clustered into Operational Taxonomic Units (OTUs) by *de novo* OTU picking in USEARCH (Edgar, 2010). Chimeras were detected and omitted with UCHIME (Edgar et al., 2011) and the QIIME-compatible version of the SILVA 111 release database (Quast et al., 2013). The most abundant sequence was chosen as the representative for each OTU. Taxonomy was assigned to representative sequences by BLAST (Altschul, 1990) against the SILVA 111 release database (Quast et al., 2013). Representative sequences were aligned against the Greengenes core reference alignment (DeSantis et al., 2006) using PyNAST (Caporaso et al., 2010a) and the phylogenetic tree was drawn with Fasttree (Price et al., 2009).

4.3.3 Transformation of *Tatumella* TA1

Plasmid pZEP-GFP (green fluorescent protein) (Hautefort et al., 2003) carrying a kanamycin-resistance cassette was electroporated into *Tatumella* TA1. Late-exponential phase cells grown in Brain Heart Infusion (BHI) Broth were pelleted at 4 ^{o}C by centrifugation at 4000 rpm for 15 minutes and washed in 30 ml ice-cold sterile distilled water. Washing was repeated twice. The cells were resuspended in 200 µl ice-cold sterile water and kept on ice

until use. 100 ng of salt-free plasmid DNA was mixed with 50 µl electrocompetent cells, transferred to an ice-cold 2 mm gap cuvette, and pulsed at 2.5 kV in a BioRad Gene Pulser (BioRad). 1 ml of room temperature BHI was immediately added to the pulsed cells and the mixture was incubated at 25 ^{o}C overnight. Transformed cells were selected by plating on to BHI agar containing 50 µg ml^{-1} kanamycin and detection of GFP-fluorescence after incubation at 25 ^{o}C for 48 h. The identity of transformed TA1-GFP cells was verified by 16S rRNA gene sequencing. To prepare the GFP-TA1 suspension for inoculation experiments, single colonies of TA1-GFP were inoculated in to BHI with 50 µg ml^{-1} kanamycin and incubated in shaking (100rpm) culture at 25 ^{o}C. Cultures were grown to an OD600 of 0.3-0.4, pelleted at 4000 x g for 10 minutes at room temperature and re-suspended in sterile water to a concentration of 1 x 10^{8} cells ml^{-1} with 50 µg ml^{-1} kanamycin.

4.3.4 TA1-GFP inoculation

Juvenile inoculation began immediately upon egg collection, and all inoculations were performed in triplicate. There were three treatments in total: TA1-GFP, water, and water with 50 µg ml^{-1} kanamycin (referred to as 'control suspension'). 1 ml of TA1-GFP suspension or control suspension was added directly to the filter paper in each petri dish of eggs prior to their application (~2000 eggs) and incubated at 25 ^{o}C for 48 h. Upon hatching, 200 first instar larvae were transferred to larval diet with 2 ml of TA1-GFP suspension or control solution (n=10 dishes of 200 larvae per treatment). To facilitate the addition of Gram-negative bacteria to the larval diet, the hydrochloric acid that is used to prevent the growth of bacteria was excluded (Hagen et al., 1963). Dishes were incubated at 25 ^{o}C and RH 60 %. After each larval moult (on days 4 and 8 respectively) a four-times concentrated TA1-GFP suspension or control solution was added to the diet. The suspension was concentrated in order to reduce the amount of liquid applied to the diet. Larvae were allowed to develop to pupation.

For adult inoculations, pupae were placed in to cages in groups of 50 for longevity experiments (n = 3 per treatment) and 100 for locomotor

experiments (n = 1 per treatment) and allowed to emerge over a 48 h period. For each 50 flies, 2 ml of TA1-GFP or control suspension was supplied in drinking water by soaking cotton wool feeders in solution (i.e. cages with 100 individuals received double the amount of suspension). Inoculations began 48 h after the emergence of the first adults and were repeated every 24 h for 72 h. 24 h after each inoculation guts were dissected from three surface-sterilized (washed in 70 % ethanol and rinsed in sterile water) individuals per treatment. Individual guts were homogenized in 200 µl 1 x PBS with a plastic pestle: 20 µl of lysate were plated on to nutrient agar, 20 µl on to nutrient agar with 50 µg ml^{-1} kanamycin. 72 h after the first inoculation single insects from all three treatments were placed in to 7 mm glass tubes containing 2 % agar and 4 % sucrose diet and were monitored from days 7 to 14 after emergence.

4.3.5 Tetracycline treatment

Adult wild and Argov strain flies were treated with tetracycline to remove their microbiota. Flies were allowed to emerge from their puparium and feed on conventional diet. Within 48 h of emergence, single insects were placed in to 7 mm glass tubes containing 2 % agar and 4 % sucrose diet with or without 100 µg ml^{-1} tetracycline and were monitored from days 2 to 7 after emergence (Table 4.1).

4.3.6 Activity monitoring

Tubes were randomly placed in to six Drosophila Activity Monitor units (Pfeiffenberger et al., 2010) along with empty tubes as negative controls, and all units were placed in to an incubator maintained at 25 ^{o}C with a 12:12 h light:dark cycle identical to that in place during rearing. Data were continuously logged in 5-minute bins. Insect metadata for each experiment is outlined in Table 1. Activity monitoring data was error-checked with DAMFileScan version 110X (www.trikinetics.com) and analysed in R version 3.2.3 (R Core Team, 2015) using custom scripts. Inactivity for a period of 24 h was used as a proxy for death.

Table 4.1 | Metadata for individual insects subjected to activity monitoring. *Days post-emergence.

Strain	Treatment	No. Females	No. Males	Total No. Insects	Dpe* Day 1 Experiment	No. Days Monitoring
Crete	Normal	64	63	127	2	5
Crete	Tetracycline	16	16	32	3	5
Argov	Normal	24	40	64	2	5
Argov	Tetracycline	16	16	32	3	5
Argov	Normal	20	33	53	7	8
Argov	Kanamycin	27	18	45	7	8
Argov	Kanamycin-TA1-GFP	20	27	47	7	8

4.3.7 Adult longevity

An independent measure of longevity was conducted in order to validate longevity estimates from activity monitoring data for flies reared normally or treated with kanamycin and TA1-GFP. Three cages of 50 individuals were established per treatment as above. The number of dead individuals per cage was counted each day from the first day of inoculations (one-day post-emergence) for a period of 14 days, and mortality curves were drawn in R version 3.2.3.

4.4 Results

4.4.1 Ontogenetic variation in the bacterial communities

16S rRNA gene PCR and digestion with PstI showed that wild larvae were dominated by *Ca.* E. dacicola, but also housed other bacteria (Figure 4.1). 16S rRNA gene amplicon profiling of Argov strain *B. oleae* showed variation in community structure and composition over the course of development (Figure 4.2). Eggs and adults shared a microbiota that was dominated by *Proteobacteria*, and was very similar in composition and structure at the phylum level. However, this condition was not maintained in larvae or pupae. In these communities the proportion of *Proteobacteria* decreased with each larval moult in to pupation, and was replaced by an increasing proportion of Gram-positive *Firmicutes*.

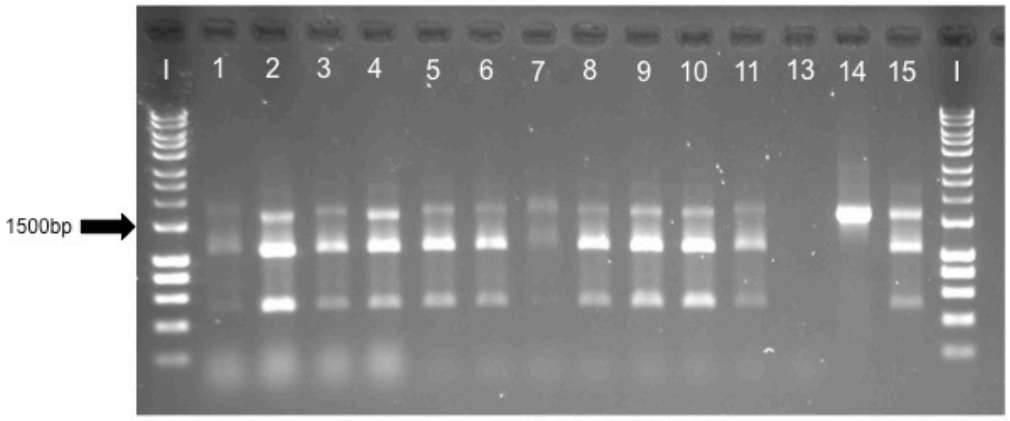


Figure 4.1 | Restriction digest of 16S rRNA PCR amplicons generated from wild larval DNA with PstI. Lanes marked with I are HyperLadder I (BioLine). Lanes 1 to 11 are wild larvae collected in autumn 2015. Lane 13 is a negative control, lane 14 is *Arsenophonus* DNA and lane 15 is a *B. oleae* adult sample with *Ca.* E. dacicola at 85 % relative abundance. Restriction digest indicates that the bacterial communities in wild larvae are similar in structure and composition to those in wild adults, and are dominated by *Ca.* E. dacicola.

4.4.2 Inoculation of juvenile laboratory-reared flies with TA1-GFP

Removal of the hydrochloric acid that acts as a preservative in standard larval artificial diet for *B. oleae* resulted in 100 % mortality over the course of juvenile development, irrespective of kanamycin or GFP-TA1 treatment. No pupae emerged from the 6000 first instar larvae that were seeded on to treated larval diet (n = 2000 per treatment). The larval diets were discoloured

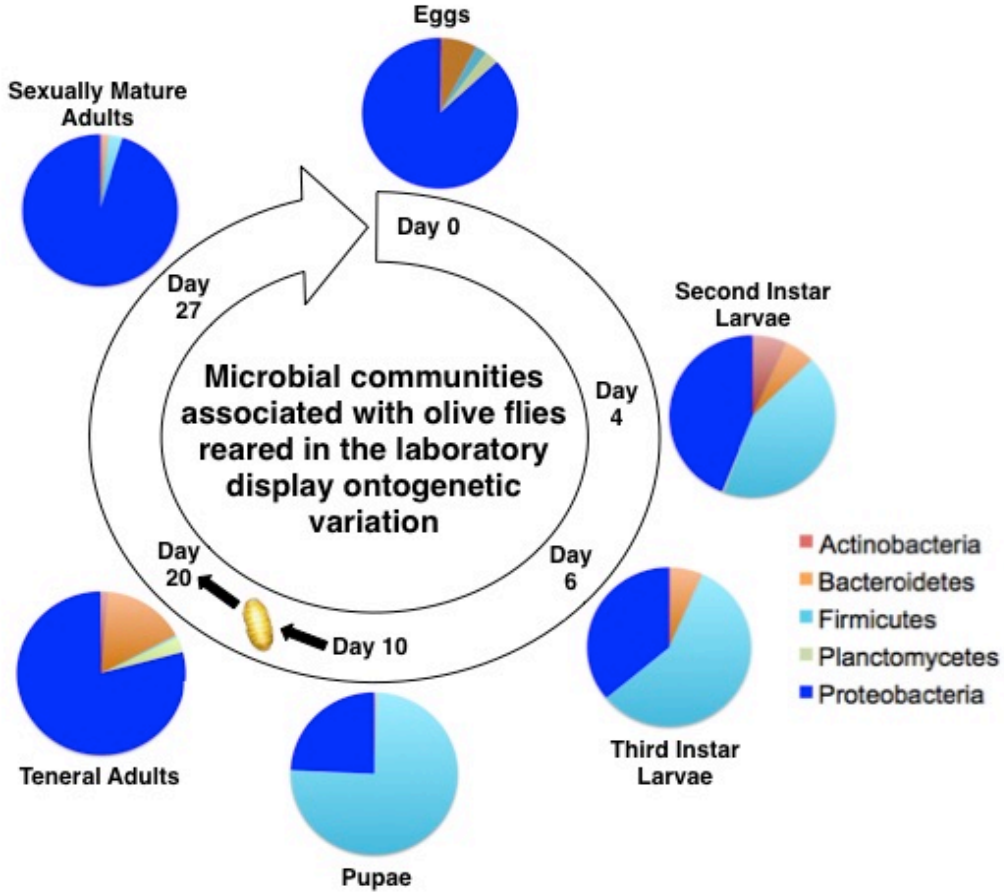


Figure 4.2 | Variation in the relative abundance of bacterial phyla during the *B. oleae* life cycle in Argov flies measured by 16S rRNA gene amplicon profiling of the V4 region on an Illumina MiSeq sequencer. Profiles indicate that juveniles and adults do not have the same bacterial community composition and structure.

and foul smelling, and though the etiological agent was not identified it was presumed to be of microbiological origin.

4.4.3 The effect of tetracycline treatment on activity and longevity

The locomotion of Argov laboratory-reared flies and wild flies from Crete was measured using the Drosophila Activity Monitoring (DAM) system. Tetracycline treatment induced sex and strain-specific differences in survival, mean activity, and somnolence. Mortality was significantly higher in Argov flies than in wild flies from Crete, regardless of antibiotic treatment (Figure 4.3), and was highest in Argov females treated with tetracycline. This is in contrast to Argov males treated with tetracycline, where survival was higher than in untreated individuals, and to wild insects where tetracycline treatment did not significantly affect mortality in either sex. Tetracycline treatment

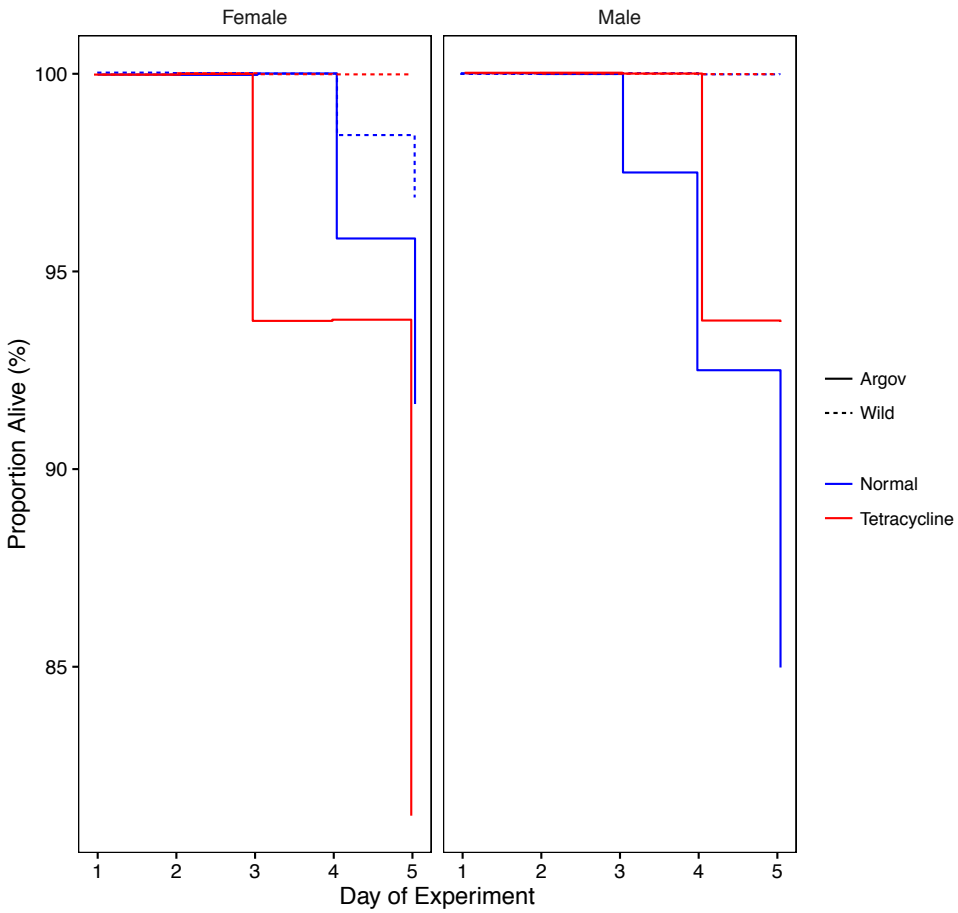


Figure 4.3 | Survival of wild and Argov strain adults treated with tetracycline assessed by activity monitoring over a period of five days.

influenced activity in a sex-specific manner in both fly strains. Argov males were unaffected by antibiotic treatment, but Argov females spent less time awake and were less active during their waking periods when treated with tetracycline. Conversely, wild females were unaffected by antibiotic treatment, but wild males displayed reduced mean activity and increased somnolence under tetracycline treatment (Figure 4.4A and B).

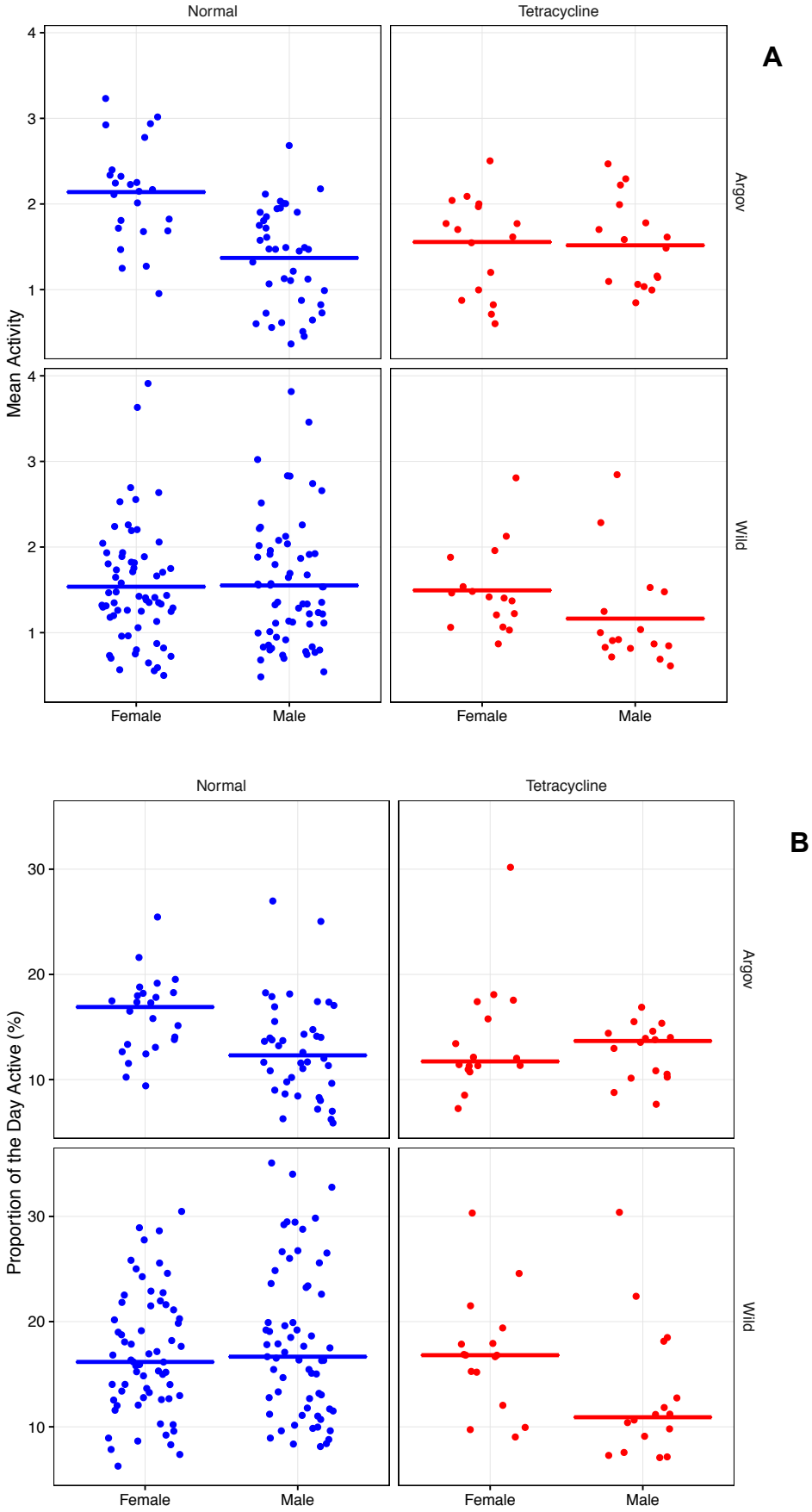


Figure 4.4 | **A** Mean activity and **B** Proportion of the day active of wild and Argov adults treated with tetracycline. Points represent individual averages and bars represent the treatment group mean over the course of the experiment.

Table 4.1 | Bacteria isolated from Argov strain *B. oleae* treated with kanamycin and TA1-GFP from the same cohort of flies that were subjected to activity monitoring.

Treatment	Mean CFUs per Gut	Genus
Water	2×10^7	*Serratia*
Kanamycin	3×10^3	*Stenotrophomonas*
Kanamycin and TA1-GFP	1×10^8	*Stenotrophomonas*

4.4.4 Experimental perturbation of the microbiota

A preliminary assessment of the impact of kanamycin and TA1-GFP treatment on the microbiota was carried out by enumerating the culture-viable members of the community by dilution plating on the last day of activity monitoring (Table 4.1). CFU counts indicated that adult flies were not successfully inoculated with TA1-GFP, but that both treatments induced changes in the microbiota. The burden of bacteria was equivalent between conventionally reared and kanamycin-TA1-GFP treated flies, but that the most dominant member of the community differed between the two treatments. In conventionally reared flies *Serratia* was dominant, as was found previously by 16S rRNA gene amplicon profiling of Argov flies from this colony (Chapter 3), whereas in kanamycin-TA1-GFP treated flies the genus *Stenotrophomonas* dominated the community. Flies treated with kanamycin only were also dominated by *Stenotrophomonas* bacteria, but at much lower titre than in the other two treatments.

4.4.5 Activity and longevity after microbiota perturbation

The locomotion of Argov strain flies treated with kanamycin and TA1-GFP was measured using the DAM system. On average, and independent of treatment, mortality was higher in males than in females (Figure 4.5). The exception to this trend was the female cohort treated with kanamycin and TA1-GFP, where mortality was significantly higher than in females reared conventionally or on kanamycin only. The variation in mortality between female treatment groups was not reflected in either mean activity over the

course of the experiment (Figure 4.6) or in the proportion of the day spent active (Figure 4.7). In males, however, the inverse was true: both activity and somnolence varied with treatment, whereas survival was not affected. While kanamycin treatment alone did not alter mean activity, the addition of TA1-GFP caused a significant decrease in mean activity after day 5 of monitoring in males (Figure 4.6). Kanamycin treatment both with and without GFP-TA1 increased somnolence in comparison to males reared conventionally (Figure 4.7).

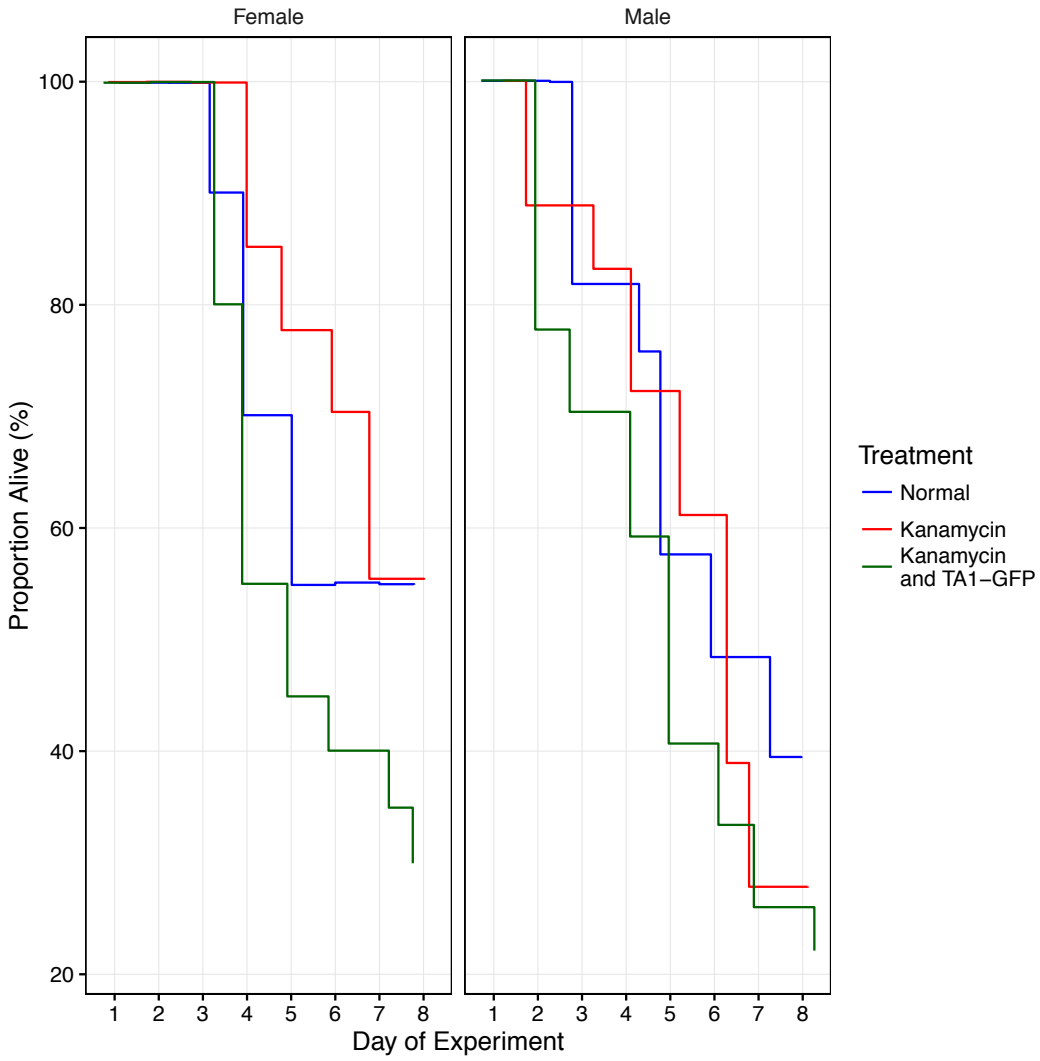


Figure 4.5 | Survival of adults with experimentally altered gut microbiota. Adults were inoculated with water (Normal), kanamycin, or kanamycin and TA1-GFP. Activity was monitored over a period of eight days.

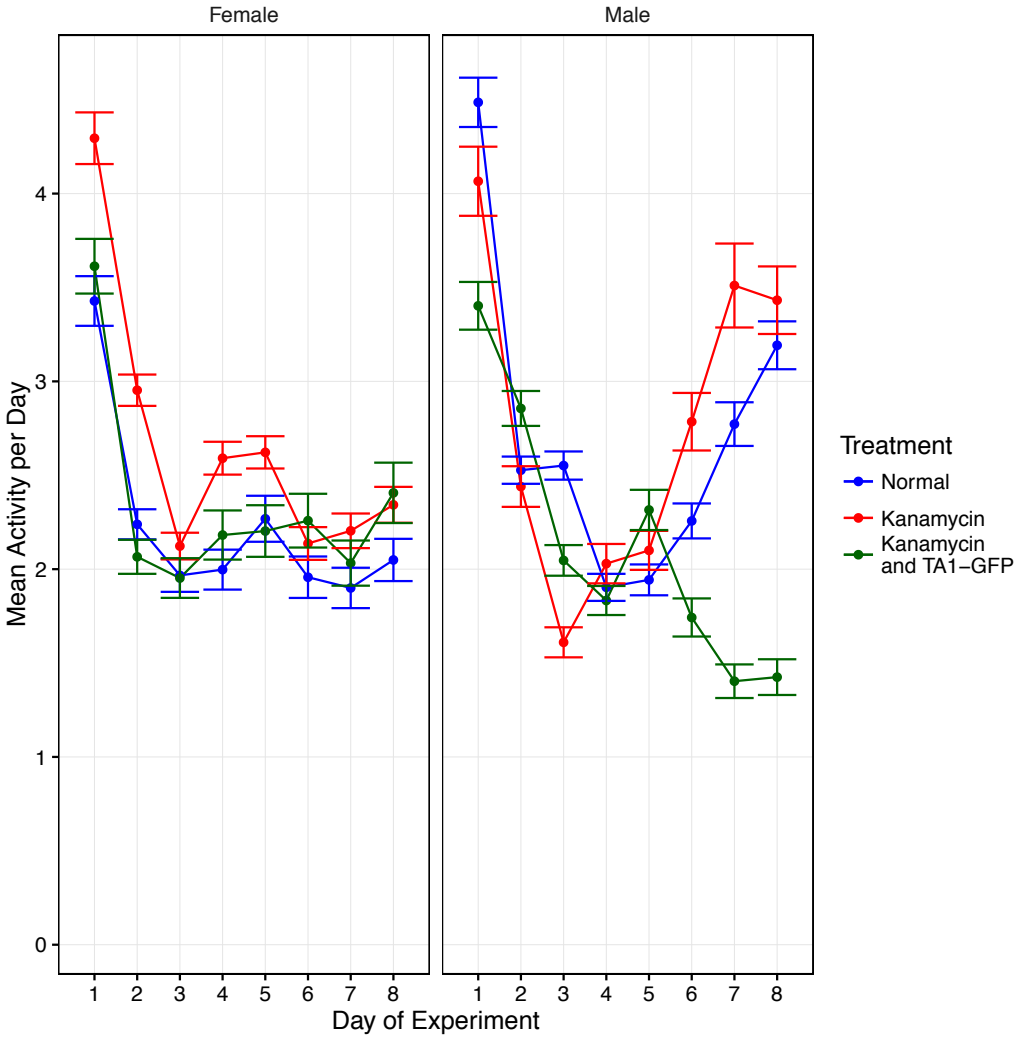


Figure 4.6 | Mean activity per treatment group per day of Argov *B. oleae* adults with experimentally altered microbiota. Points represent the mean activity per treatment group per day and error bars are the standard error of the mean.

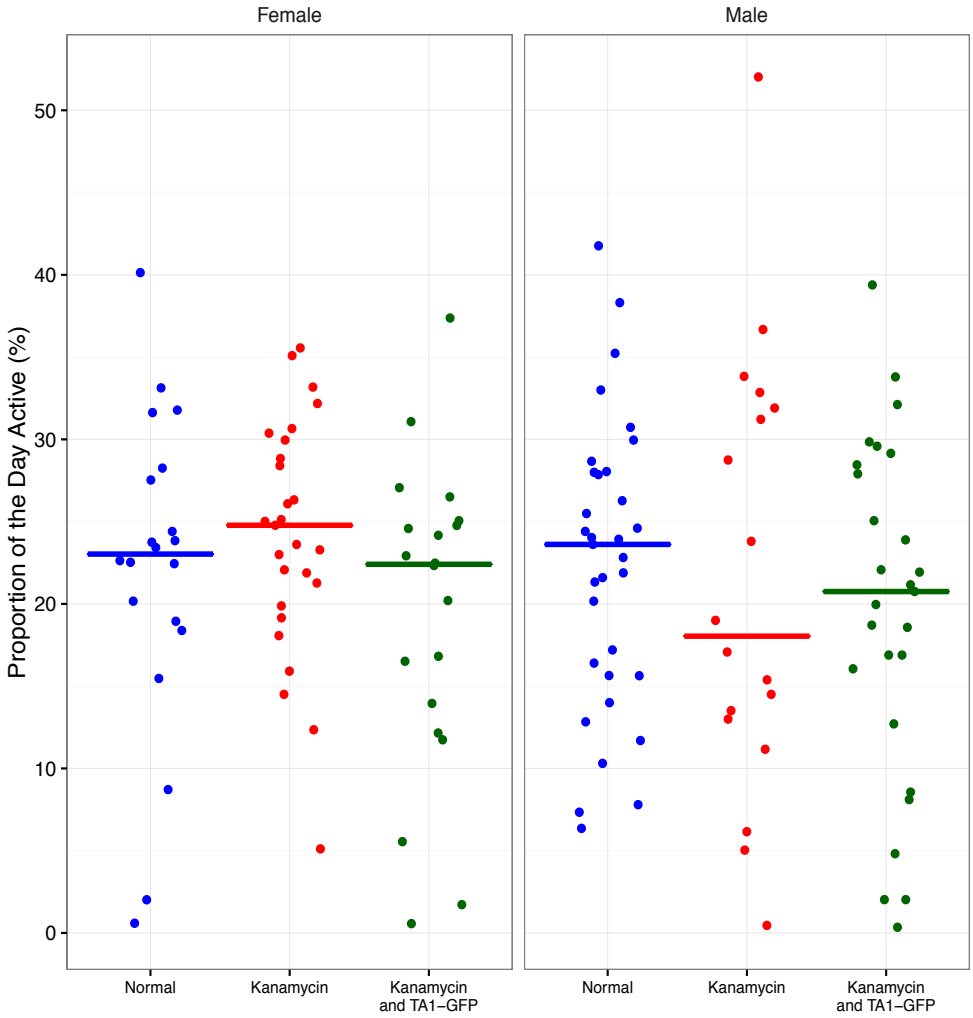
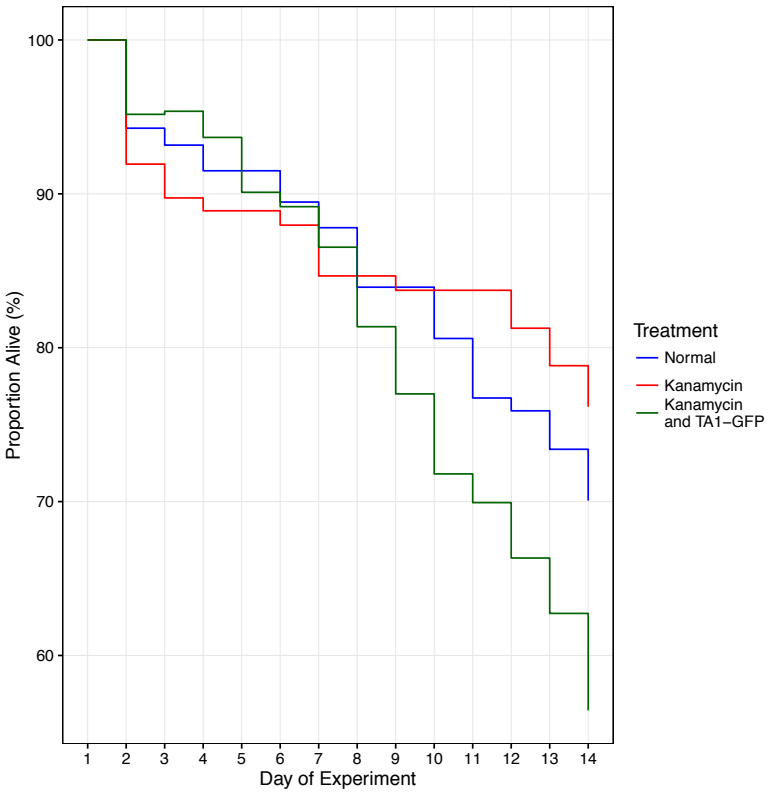


Figure 4.7 | The proportion of the day spent active by male and female Argov *B. oleae* with experimentally altered microbiota. Points represent the average time spent awake by each individual over the course of the eight-day experiment. Bars indicate the median time spent awake per treatment group.

Longevity was measured independently of activity monitoring by counting the number of dead individuals per day. Sex-specific mortality was not measured during this experiment (Figure 4.8A), so was compared to total mortality per treatment for activity-monitored flies (Figure 4.8B). Trends were similar between experiments: mortality was higher in kanamycin-GFP-TA1-treated individuals than in those reared conventionally or on kanamycin only. In the conventional longevity experiment mortality was higher in flies reared normally than in those reared on kanamycin, whereas the difference in mortality was negligible between these treatments during activity monitoring.



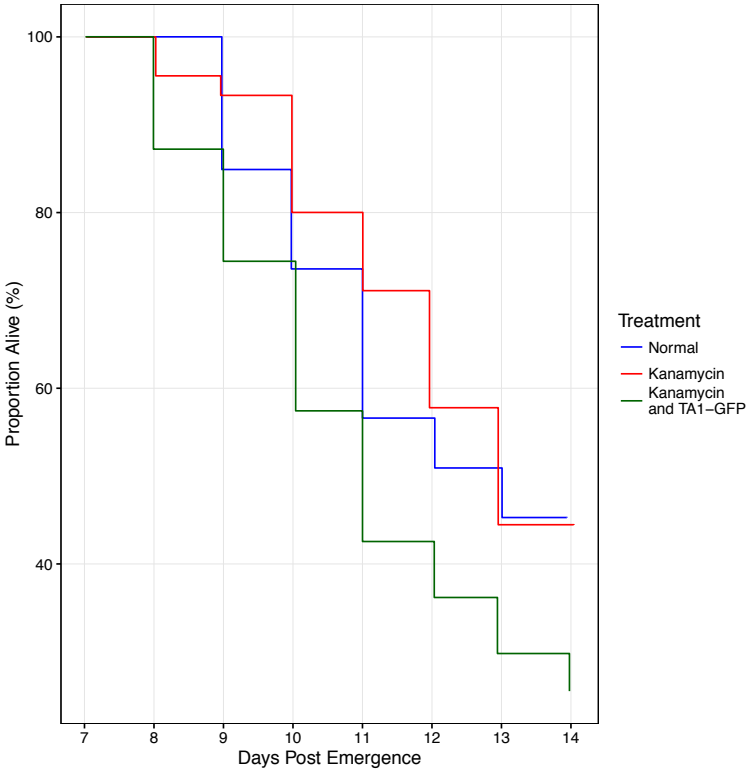


Figure 4.8 | Survival curves of Argov adults with experimentally altered microbiota measured by **A** conventional longevity experiments and **B** activity monitoring, where individuals were considered dead after a 24 h period of inactivity.

4.5 Discussion

16S rRNA gene amplicon profiling studies of the Argov laboratory strain indicate ontogenetic variation in the structure and composition of the microbiota throughout the life cycle. An extreme shift in composition was observed between the juvenile and adult stage, which correlates with a change in pH between diets. The pH of the larval diet is low to prevent the growth of pathogenic microorganisms (Hagen et al., 1963). However, low pH can lyse Gram-negative cells, which constitute the majority of the microbiota at this stage of development in natural *B. oleae* populations (Estes et al., 2012a; Ben-Yosef et al., 2014; 2015). Therefore, it is hypothesised that including HCl in the larval diet, which successfully deters the growth of pathogens, also alters the physiological conditions of the *B. oleae* gut or the diet such that commensal organisms are less successful. Instead, there is a bias towards Gram-positive bacteria such as Lactobacilli, which are better able to withstand low pH (Cotter and Hill, 2003). There is little evidence of taxonomically similar bacteria in wild populations, and no evidence of similar taxa at equivalent relative abundances (Estes et al., 2012b). However, the wild microbiota provides no nutritional advantage in a laboratory setting, which is thought to be the main selective pressure shaping microbial communities in wild *B. oleae* populations (Hagen, 1966). Therefore, the taxa that successfully replace the native microbiota at the larval stage in the laboratory may serve extraneous functions to nutrition, or are simply the best candidates from the environment at colonizing the niche (Broderick, 2016). Due to the methodology used here there is no evidence that the microbes recovered are able to stably colonize within *B. oleae*, or in fact if they are even viable. They may simply be an alternative food source, or present as a by-product of insects feeding on non-sterile diet.

Despite the inhospitable conditions of the larval diet, some Gram-negative taxa are able to colonize embryos and infect all subsequent life stages. These taxa could be vertically transmitted and maintained throughout metamorphosis, or reacquired from the environment at each stage. The

symbiont *Ca.* E. dacicola is beneficial at every life stage under natural conditions and *B. oleae* has evolved methods to ensure its efficient transferral between life stages and generations (Estes et al., 2009). It is possible that these mechanisms outlive the presence of the symbiont in laboratory populations, and that other microbes exploit vestigial transmission features for their own gain (Salem et al., 2015a). It is also known that tephritids have a predisposition towards housing *Enterobacteriaceae* (Lauzon, 2003), and therefore that the host environment, including its immune response, may be altered in some way to permit the presence and persistence of these taxa. Excluding HCl from the larval diet and thus raising the pH led to severe contamination and 100 % larval mortality. Therefore, altering the larval diet to be more receptive to Gram-negative taxa presumed to be more similar to those found in natural populations is not a viable strategy to improve rearing efficiency.

Bacterial inoculation, antibiotic treatment, and activity monitoring experiments provide further evidence that the *B. oleae* microbiota is unstable. Behavioural responses to antibiotic treatment were sexually dimorphic and depended upon host genetic background. They were also specific to the antibiotic used, even though tetracycline and kanamycin function by the same mode of action (Chopra and Roberts, 2001). This may have been due to the administration of antibiotics at different concentrations: tetracycline was supplied at 100 μg ml^{-1} and kanamycin at 50 μg ml^{-1}. The Bateman principle posits that immune responses should be sexually dimorphic, as males and females adopt different strategies to optimise their fitness (Rolff, 2002). Females increase their fitness by maximising longevity and therefore providing more time for reproduction, whereas males gain fitness by increasing their mating success (Bateman, 1948). These alternative strategies stem from the disparity in reproductive investment and genetic traits between sexes (Rolff et al., 2005). There was sexual dimorphism in response to antibiotic treatment and bacterial challenge in all of the strains and cohorts studied by activity monitoring. Argov females in particular were significantly affected by tetracycline treatment in all of the parameters measured, whereas kanamycin induced greater somnolence in Argov males. Tetracycline-based antibiotics are known to

affect insect physiology, as they also target host mitochondria (Ballard and Melvin, 2007). The disparity between tetracycline and kanamycin treatment indicates that the concentration of tetracycline required for suppression of the fsRIDL gene during rearing may incur severe fitness costs. Whether this is due to alteration of the microbiota or direct interaction of the antibiotic with host mitochondria cannot be determined from this experiment.

Argov flies were not successfully inoculated with TA1-GFP. Culturing methods indicated that TA1-GFP was not able to colonize the *B. oleae* gut at high titre over the course of the experiment. This may have been because it was supplied alongside kanamycin with a resistance plasmid, which allowed it to attenuate the environment by metabolising kanamycin, a process known as facilitation. This could either allow other microbes to thrive, or select for microbes such as *Stenotrophomonas* to acquire the resistance plasmid by Horizontal Gene Transfer (HGT).

Despite the lack of colonisation by the target organism, the experiment still provided interesting results. TA1-GFP treatment induced significant and sexually dimorphic changes in behaviour compared to flies fed on normal diet or kanamycin alone. Responses were non-uniform, but all constituted behaviours that have been negatively correlated with host health in previous studies. Females experienced increased mortality when colonised by *Stenotrophomonas* at high density (TA1-GFP treatment). However, the females that survived displayed no alterations to behaviour. Surviving females may either have resisted infection by *Stenotrophomonas*, or were better able to tolerate the infection through adaptive responses (Ayres, 2008). Alternatively, males experienced no greater mortality under treatment with TA1-GFP, due to their naturally higher mortality rates than females, but did display behavioural alterations through increased somnolence and lethargy. These alterations were presumably adopted to mitigate the costs of infection by re-allocating resources (Schmid-Hempel, 2005). The same phenomenon was observed in male *D. melanogaster* infected with *w*Mel *Wolbachia*, indicating that it could be a general response to bacterial infection (Vale and Jardine, 2015). These results fit with Bateman's principle (Rolff, 2002). In

some females, this tactic may have resulted in increased mortality due to the high metabolic cost of immune responses (Ardia et al., 2012). However, some showed no response to infection in the parameters measured, reflecting the potential for within-population genetic variation in both resistance and tolerance to infection (Råberg et al., 2007; Howick and Lazzaro, 2014; Kutzer and Armitage, 2016).

Treatment with TA1-GFP and kanamycin both resulted in removal of the *Serratia* commonly associated with Argov flies and its replacement with *Stenotrophomonas*, though at different densities. Although *Serratia* is never found as a dominant member of the wild microbiota, it was the most abundant bacterium in Oxitec Ltd. laboratory-reared flies (Chapter 2), indicating that the association is consistent between generations in this rearing environment. There is also evidence that it is either vertically transmitted and maintained throughout metamorphosis, or reacquired at every life stage, including the embryonic stage. Given this evidence, and that flies displayed increased mortality or sickness behaviours when *Stenotrophomonas* replaced *Serratia* at a similar density, it is possible that *Serratia* is a commensal bacterium of laboratory-reared *B. oleae*.

Preliminary inoculation experiments performed with antibiotics and TA1-GFP raise several questions about the mechanisms by which *B. oleae* interacts with both commensal and pathogenic organisms, particularly in a mass-rearing environment. Discovery of a potentially commensal *Serratia* sp., many of which form symbiotic and pathogenic interactions with insects (Grimont and Grimont, 1978; Moran et al., 2005), suggests that there may be a low-diversity, lab-adapted microbiome in *B. oleae*. In order to better understand this relationship, further studies are required to determine its nature. For example, does *Serratia* provide any functional benefit to the host such as protection by direct antagonism with opportunistic pathogens or niche occupation, or is it somehow involved in nutrition, as the wild microbiota is? Alternatively, is it an opportunistic microbe that is simply tolerated by the host? Due to limitations in experimental design, the function of the hypothesised mutualist TA1 also remains elusive, as do the infection

dynamics of laboratory-reared flies when treated with antibiotics. Future work should investigate whether TA1-GFP is able to colonize the *B. oleae* gut, its residence time, localisation in relation to other members of the gut microbiota, and how this interacts with total bacterial density and fly health.

Chapter 5 : Challenging the *Wigglesworthia*, *Sodalis*, *Wolbachia* symbiosis dogma in tsetse flies: *Spiroplasma* is present in natural and laboratory populations

I performed the data analysis for all 16S rRNA gene amplicon datasets, and wrote this chapter in collaboration with several co-authors.

Doudoumis V*, Blow F*, Saridaki A, Augustinos A, Dyer N, Goodhead IB, Solano P, Rayaisse JB, Takac P, Parker AG, Abd-Alla AMM, Darby AC, Bourtzis K, Tsiamis G. (In preparation). Challenging the *Wigglesworthia*, *Sodalis*, *Wolbachia* symbiosis dogma in tsetse flies: *Spiroplasma* is present in both laboratory and natural populations.

*These authors contributed equally

5.1 Abstract

Background: Tsetse flies transmit African trypanosomes, which cause sleeping sickness in humans and nagana in animals. Tsetse are considered to carry three vertically transmitted endosymbionts: *Wigglesworthia glossinidia*, *Sodalis glossinidius*, and *Wolbachia pipientis*. We investigated the distribution and relative abundance of bacterial symbionts in genetically distinct populations of tsetse flies from natural and laboratory environments in order to identify the influence of host genotype and ecology on symbiont infection patterns.

Methods: We used 16S rRNA gene amplicon sequencing to assess whether the "*Wigglesworthia Sodalis Wolbachia* dogma" is common across species and populations. PCR screening, qPCR, whole genome shotgun sequencing, MLST and Fluorescent In Situ Hybridization (FISH) were subsequently used to validate the results of 16S rRNA gene amplicon sequencing.

Results: 16S rRNA gene amplicon profiling of wild tsetse populations revealed variation in bacterial community composition and structure between tsetse species. Several novel taxa were detected in addition to the three previously characterized symbionts. These included *Klebsiella*, *Rickettsia* and *Spiroplasma*, all of which have previously been found in arthropods. *Spiroplasma* was detected in multiple individuals from two tsetse species: *Glossina fuscipes fuscipes* and *G. tachinoides*, both of with belong to the *Palpalis* subgroup. MLST analysis indicated that *G. f. fuscipes* and *G. tachinoides* house distinct *Spiroplasma* strains, and FISH localized the Spiroplasma strain from *G. f. fuscipes* to the reproductive tissues of males and females.

Conclusions: The bacterial communities associated with *Glossina* are not restricted to *Wigglesworthia*, *Sodalis* and *Wolbachia*, and include taxa that have been characterised as mutualists and pathogens in other insects. The genomic resources generated during this study will enable future investigations in to the diverse range of microbes associated with tsetse flies and their potential impact on vector biology.

5.2 Introduction

The microbiota of tsetse flies is of interest because of their unique lifestyle and reproductive strategy as well as its potential for vector and disease control (Welburn et al., 2001; Aksoy, 2011; Abd-Alla et al., 2013; Wang et al., 2013; Bourtzis et al., 2016). So far, it is known that tsetse flies harbour three main symbiotic microbes: *Wigglesworthia*, *Sodalis* and *Wolbachia*. The primary mutualist *Wigglesworthia glossinidia* provides dietary supplements that are necessary for host fecundity, larval development, and the maturation process of the adult immune system (Aksoy, 1995; Weiss et al., 2011; Weiss and Aksoy, 2011; Weiss et al., 2012). The facultative symbiont *Sodalis* is present in tsetse populations with a putative role in the ability to transmit trypanosomes (Farikou et al., 2010). *Sodalis* utilizes a vertical transmission route, similar to that of *Wigglesworthia*, and can be horizontally transmitted during mating, resulting in paternal vertical transmission (De Vooght et al., 2015). Finally, *Wolbachia* has been found in natural populations of tsetse flies with some species exhibiting up to 100 % infection rate (Alam et al., 2011). The *Wolbachia* strain present in *Glossina morsitans morsitans* (*Gmm*) can induce cytoplasmic incompatibility under laboratory conditions (Alam et al., 2011). This property, together with the ability of *Wolbachia* to provide protection against pathogens and parasites, has sparked interest in its potential for tsetse vector and disease control (Xi et al., 2005; Alam et al., 2011; Doudoumis et al., 2013; McGraw and O'Neill, 2013; Bourtzis et al., 2014; Geiger et al., 2015; Bourtzis et al., 2016).

There have been a limited number of culture-dependent and culture-independent studies aiming to characterize the microbiota associated with tsetse flies. Using classical microbiological approaches, Geiger and colleagues isolated *Acinetobacter, Enterobacter, Enterococcus*, *Providencia*, *Sphingobacterium*, *Chryseobacterium*, *Lactococcus*, *Staphylococcus* and *Pseudomonas* species from the guts of field collected *G. p palpalis* and less abundant species in Cameroon (Geiger et al., 2009) (Geiger et al., 2010; 2011; 2013). They also isolated a new bacterial species, *Serratia glossinae*,

from the midgut of *G. palpalis gambiensis* collected in Burkina Faso (Geiger et al., 2010). Lindh and Lehane (Lindh and Lehane, 2011) performed a screen for both cultivable and non-cultivable bacteria occurring in whole *G. fuscipes fuscipes* flies collected in Kenya. *Firmicutes*, and particularly members of the *Bacillus* genus, were identified as the most dominant groups while *Paenibacillus*, *Staphylococcus* and *Exiguobacterium* species were also isolated at lower density. *Gammaproteobacteria* were also present, mainly members of the *Enterobacteriaceae* like *Morganella* and *Providencia* and, to a lesser degree, *Pseudomonas* spp. *Burkholderia* was the only member of the *Betaproteobacteria* detected in this study (Lindh and Lehane, 2011). Interestingly, the sensitivity of the non-culture viable DGGE (Denaturing Gradient Gel Electrophoresis) screen was insufficient to uncover many associated bacterial taxa. Beyond the mutualist symbiont *Wigglesworthia,* only *Bacillus* and *Serratia* spp. were detected (Lindh and Lehane, 2011). Aksoy et al (Aksoy et al., 2014) sampled guts of Ugandan *G. f. fuscipes*, *G. m. morsitans* and *G. pallidipes* tsetse flies, and profiled the microbiota using Illumina MiSeq sequencing of the V4 variable region of the 16S rRNA gene. *Wigglesworthia* was the dominant taxon, while *Sodalis* was generally detected at low density (<0.05 %). However, a small number of flies harboured high levels of *Sodalis* and *Serratia* spp. *Halomonas* spp. and non-*Wigglesworthia Enterobacteriaceae* were also found at lower abundance at all field sites studied, with some bacterial taxa being unique to a sample site.

Spiroplasma is a genus of wall-less bacteria belonging to the class *Mollicutes.* Members of this genus have been associated with diverse plants and arthropods (Whitcomb et al., 1986). *Spiroplasma* is grouped in to three major clades as has been shown by 16S rRNA gene-based as well as multi locus sequence typing (MLST) studies (Anbutsu and Fukatsu, 2003; Montenegro et al., 2005; Bi et al., 2008; Haselkorn et al., 2009; Heres and Lightner, 2010; Haselkorn et al., 2013; Nai et al., 2014). *Spiroplasma* is able to live both intracellularly in a variety of tissues and systemically in the haemolymph (Schwarz et al., 2014). *Spiroplasma* spp. have developed a wide range of symbiotic associations, producing diverse effects on insect evolution, ecology, reproduction and sex determination. They confer protection against a

nematode in *Drosophila neotestacea* (Jaenike et al., 2010), against fungi in the pea aphid (*Acyrthosiphon pisum*) (Lukasik et al., 2013), and against a parasitoid wasp in *Drosophila hydei* (Xie et al., 2013). They can also be pathogenic in plants (Gasparich, 2010), insects (Mouches et al., 1984; Clark et al., 1985; Humphery-Smith et al., 1991) and crustaceans (Wang, 2004; Nunan, 2005; Wang et al., 2005; Liang et al., 2011; Wang et al., 2011; Ding et al., 2012). Moreover, several species of *Spiroplasma* have been associated with reproductive alterations such as male killing (Williamson et al., 1999; Hurst, 2000; Anbutsu and Fukatsu, 2003; Tabata et al., 2011; Hayashi et al., 2016). Interestingly, Spiroplasmas do not activate an immune response and are not susceptible to either the cellular or humoral arms of the immune system in *Drosophila* (Hurst et al., 2003; Herren and Lemaitre, 2011).

In this study we employed high throughput sequencing of the V4 region of the 16S rRNA gene to assess the diversity of tsetse associated bacteria in a wider variety of species, field, and lab populations than any previous tsetse microbiota study. We asked whether the "*Wigglesworthia Sodalis Wolbachia* dogma" applies across species and populations, and whether the microbiota varies between lab and field individuals of the same tsetse species. *Spiroplasma* was identified as a novel symbiont of *G. f. fuscipes* and *G. tachinoides*, and infection prevalence was surveyed in lab and natural populations. A quantitative PCR approach was used to characterize *Spiroplasma* density in different developmental stages and tissues, and to quantify infection levels in collapsing mass-reared tsetse fly colonies. FISH was used to localize the newly identified symbiont to tissues including the gonads.

5.3 Methods

5.3.1 Insect collection and DNA isolation

All natural populations of *Glossina* specimens were collected in four countries: Burkina Faso, Uganda, Tanzania, and South Africa (Table 5.1 and Table 5.3). Upon arrival in the lab, DNA was extracted immediately using the CTAB (Cetyl trimethylammonium bromide) method and was stored at -20 ^{o}C. Laboratory populations from the Joint FAO/IAEA Insect Pest Control Laboratory (Seibersdorf), Laboratory of Epidemiology and Public Health (LEPH, Yale, US), Kenya Agriculture and Livestock Research Institute-Biotechnology Research Institute (KALRO – BRI in Nairobi, Kenya; former KARI-TRC), Centre international de recherche-développement sur l'elevage en zone subhumide (CIRDES in Bobo Dioulasso, Burkina Faso), Institute of Tropical Medicine (Antwerp, Belgium), Vector and Vector Borne Diseases Research Institute (Tanga, Tanzania), and the Slovak Academy of Sciences (Bratislava, Slovakia) were treated in a similar way (Table 5.1 and Table 5.3). Samples of *G. f. fuscipes* suffering high mortality were collected from the mass rearing facility in Kality, Ethiopia. Tsetse fly tissues from laboratory-reared flies (ovaries, testes and guts) were freshly dissected under sterile conditions from teneral and 15-day-old flies. Tissue samples were pooled from five individuals and genomic DNA was isolated using the Qiagen DNeasy kit (Qiagen, Valencia, CA), following the manufacturers' instructions. In total, three biological replicates were obtained for each sample based on tissue, sex, and developmental stage for all laboratory populations.

Table 5.1 | Metadata for all samples profiled by Illumina MiSeq sequencing of 16S rRNA gene amplicons of the V4 region. (^ 3 male-only samples and 3 female-only samples. * samples analysed by PCR and amplicon profiling. & population used for MLST analysis. [1]Joint FAO/IAEA Insect Pest Control Laboratory, Seibersdorf)

Species	Population type	Collection date	Location	Tissue	Number of samples	Number of individuals
G. f. fuscipes	Natural	Aksoy et al. 2014	Uganda (Busime)	Gut	13	1
	Natural	Aksoy et al. 2014	Uganda (Dokolo)	Gut	10	1
	Natural	Aksoy et al. 2014	Uganda (Kaberamaido)	Gut	25	1
	Natural	Aksoy et al. 2014	Uganda (Murchison Falls)	Gut	10	1
	Natural	Aksoy et al. 2014	Uganda (Otubio)	Gut	13	1
	Lab	2013	Lab colony[1]	Larval	3	5
	Lab	2013	Lab colony[1]	Gut 1d	6^	5
	Lab	2013	Lab colony[1]	Gut 15d	6^	5
	Lab	2013	Lab colony[1]	Reproductive 1d	6^	5
	Lab	2013	Lab colony[1]	Reproductive 15d	6^	5
G. medicorum	Natural	2010	Burkina Faso	Whole	8	1
G. m. morsitans	Natural	Aksoy et al. 2014	Uganda (Murchison Falls)	Gut	6	1
	Lab	2013	Lab colony[1]	Larval	3	5
	Lab	2013	Lab colony[1]	Gut 1d	6^	5
	Lab	2013	Lab colony[1]	Gut 15d	6^	5
	Lab	2013	Lab colony[1]	Reproductive 1d	6^	5
	Lab	2013	Lab colony[1]	Reproductive 15d	6^	5
G. m. submorsitans	Natural*	2010	Burkina Faso	Whole	8	1
G. pallidipes	Natural	Aksoy et al. 2014	Uganda (Murchison Falls)	Gut	42	1
	Lab	2013	Lab colony[1]	Larval	3	5
	Lab	2013	Lab colony[1]	Gut 1d	6^	5
	Lab	2013	Lab colony[1]	Gut 15d	6^	5
	Lab	2013	Lab colony[1]	Reproductive 1d	6^	5
	Lab	2013	Lab colony[1]	Reproductive 15d	6^	5
G. p. gambiensis	Natural	2010	Burkina Faso	Whole	8	1
G. tachinoides	Natural*[&]	2010	Burkina Faso	Whole	8	1

5.3.2 16S rRNA gene amplicon profiling

The V4 region of the 16S rRNA gene was amplified from individual wild *G. medicorum, G. m. submorsitans, G. p. gambiensis* and *G. tachinoides* flies collected in Burkina Faso. The V3-V4 region of the 16S rRNA gene was amplified from pools of tissues from larvae and adults of laboratory populations of *G. m. morsitans*, *G. f. fuscipes*, and *G. p. gambiensis* (Table 5.1). Fusion primers U341F (5'-CCTACGGGRSGCAGCAG-3') and 805R (5'-GTGCCAGCMGCCGCGGTAA-3') were used to amplify the V3-V4 region, and fusion primers F515 (5'-GTGCCAGCMGCCGCGGTAA-3') and 805R were used to amplify the V4 region of the 16S rRNA gene. 805R reverse primers contained a unique Golay barcode specific to each sample for read de-multiplexing. Each PCR reaction was carried out in a volume of 25 µl containing 1 µl of template DNA, 12.5 µl NEBNext 2x High-Fidelity Master Mix (New England Biolabs, UK), 1 µl of each primer at 3 mM concentration and 9.5 µl PCR-clean water. Cycling conditions were 30 seconds at 98 ^{o}C, 25 cycles of 98 ^{o}C for 10 seconds, 58 ^{o}C for 30 seconds and 72 ^{o}C for 30 seconds, followed by a final extension of 5 minutes at 72 ^{o}C. Amplicons were cleaned up using Ampure XP beads (Agencourt, UK), and resuspended in 30 µl PCR-clean water. Products were quantified using the Qubit dsDNA High-Sensitivity assay (Life Technologies, UK), and an Agilent Bioanalyzer High-Sensitivity DNA chip (Agilent, UK). Samples were pooled at equimolar concentrations for size-selection by Pippin-Prep (Sage Science, UK), where a size range of 350-450 bp was extracted. Negative controls were also included. Size-selected fragments underwent the same clean up and quantification steps as above prior to sequencing. Sequencing was performed at the University of Liverpool Centre for Genomic Research, and at IMGM Laboratories GmbH on an Illumina MiSeq platform using 250 bp paired-end, and 300 bp paired-end reads, respectively.

5.3.3 Data analysis 16S rRNA gene amplicon profiling

Raw sequencing reads were de-multiplexed and converted to FASTQ format using CASAVA version 1.8 (Illumina 2011). Cutadapt version 1.2.1 (Martin, 2011) was used to trim Illumina adapter sequences from FASTQ files. Reads were trimmed if 3 bp or more of the 3' end of a read matched the adapter

sequence. Sickle version 1.200 (Joshi and Fass, 2011) was used to trim reads based on quality: any reads with a window quality score of less than 20, or which were less than 10 bp long after trimming, were discarded. BayesHammer was used to correct reads based on quality (Nikolenko et al., 2013). Paired-end reads were assembled, trimmed by length and further corrected for error using PandaSeq (Masella et al., 2012). Unassembled reads and reads outside the range of 240 - 260 bp once assembled were discarded. All subsequent analyses were conducted in QIIME version 1.7.0 (Caporaso et al., 2010b). Sequences were clustered into Operational Taxonomic Units (OTUs) using USEARCH (Edgar, 2010) by de-novo OTU picking. Chimeras were detected and omitted using the program UCHIME (Edgar et al., 2011) with the QIIME-compatible version of the SILVA 111 release database (Quast et al., 2013). The most abundant sequence was chosen as the representative for each OTU. Taxonomy was assigned to representative sequences by BLAST (Altschul et al., 1990) against the SILVA 111 release database (Quast et al., 2013). Representative sequences were aligned against the Greengenes core reference alignment (DeSantis et al., 2006) using PyNAST (Caporaso et al., 2010a). Between-sample diversity was calculated using weighted Unifrac distances, and PCoA analyses were performed on the resulting distance matrix.

5.3.4 PCR screening for reproductive manipulators

Four species of *Glossina* (*G. m. morsitans*, *G. f. fuscipes*, *G. p. gambiensis*, and *G. pallidipes*) were assayed for the presence of *Spiroplasma*, *Arsenophonus*, *Cardinium*, and *Rickettsia* marker genes by PCR. An additional six species of *Glossina* (*G. austeni, G. brevipalpis, G. m. centralis, G. m. submorsitans, G. p. palpalis* and *G. tachinoides*) were screened for *Spiroplasma* only. The primer sequences used to detect each taxon along with their target genes, product sizes, and annealing temperatures are listed in Supplementary Table 1. Thermal cycling conditions were 95 °C for 5 min followed by 35 cycles of 95 °C for 30 seconds, 30 seconds at the appropriate annealing temperature for each pair of primers and the appropriate extension time for each amplicon at 72 °C (Table 7.2), and a final extension step of 72 °C for 10 minutes. PCR reactions were performed in a total volume of 20 µl,

containing 4 µl 5x reaction buffer (Promega, UK), 1.6 µl $MgCl_2$ (25 mM), 0.1 µl deoxynucleotide triphosphate mixture (25 mM each), 0.5 µl of each primer (25 µM), 0.1µl of Taq (Promega 1U μl^{-1}), 12.2 µl PCR-clean water and 1 µl of template gDNA. All host DNA samples were positive for PCR amplification using the 12S rRNA gene arthropod universal primers, indicating satisfactory DNA quality.

5.3.5 *Spiroplasma* Multi-Locus Sequence Typing (MLST)

MLST was performed for five marker genes (*rpoB, parE, dnaA, ftsZ* and *fruR*) and a 4,702 bp region spanning the 16S rRNA, 23S rRNA and 5S rRNA genes. The primers and thermal cycling conditions used are presented in Supplementary Table 1. Briefly, 5 minutes of denaturation at 95 °C preceded thirty-five cycles of 30 seconds at 95 °C, 30 seconds at the appropriate temperature for each pair of primers and the appropriate extension time for each amplicon at 72 °C, followed by a final extension step at 72 °C for 10 minutes. PCR reactions were carried out in a total volume of 25 µl containing 12.5 µl High Fidelity Ready Mix Reaction Buffer (Promega, UK), 0.3 µl of each primer (25 µM), 10.7 µl water and 1.2 µl of template DNA. The PCR products were purified by a PEG (Polyethylene glycol) - NaCl method as described previously (Tsiamis *et al*. 2012). Both strands were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (PE Applied Biosystems) in an ABI PRISM 3710 Genetic Analyzer (PE Applied Biosystems, UK).

All nucleotide sequences were manually edited with Geneious 7.1.2. Multiple alignments were generated with MUSCLE (Edgar, 2004) and ClustalW (Thompson et al, 1994) and adjusted by eye. Phylogenetic analyses were conducted for all regions by two methods: Bayesian Inference (BI) and Maximum Likelihood. Best-fit models of nucleotide substitution for constructing phylogenies of our data sets were estimated using the Model Test and Akaike Information Criterion (Akaike, 1971) in PAUP version 4.0 (Wilgenbusch and Swofford, 2002). For six out of seven analysed *Spiroplasma* alignments (16S rRNA*,* region 16S-23S-5S rRNA*, rpoB, dnaA, parE,* and *ftsZ*), the submodel GTR+I+G was selected, while the model K81uf+I was selected for the gene *fruR*. Bootstrap values were obtained with

1,000 replications. Maximum-Likelihood trees were constructed with MEGA version 5.2.2 while Bayesian phylogenetic analyses were performed using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003) as implemented in Geneious 7.1.2 (Kearse et al., 2012). Run parameters included four Markov chain Monte Carlo (MCMC) chains with one million generations, sampled every 100 generations, and with the first 10,000 trees discarded as burn-in. Posterior probabilities were computed for the remaining trees.

5.3.6 qPCR analysis of *Spiroplasma* density

Spiroplasma density was quantified by qPCR using the *dnaA Spiroplasma* specific primers FqdnaA/RqdnaADoud for 35 cycles at 56 °C and normalized to the host *β-tubulin* gene, which was amplified using primer set GmmtubqF/GmmtubqR for 35 cycles at 56 °C. The primers used for the qPCR experiments are presented in Supplementary Table 1. Reactions were performed in duplicate with the KAPA SYBR FAST qPCR Kit (KAPA Biosystems, UK) in a total volume of 10 µl containing: 5 µl 2x KAPA SYBR FAST qPCR Master Mix Universal, 0.08 µl of each primer (25 µM), 4.34 µl water and 0.5 µl total DNA (~ 20 ng). Real-time PCR runs were conducted using MJ Research Opticon 2 (MJ Research, USA). Amplifications were carried out in a 96-well plate and each biological sample was run in triplicate. Symbiont density was defined as the copy number of the *dnaA Spiroplasma* specific fragment relative to the copy number of the *β-tubulin* host fragment (Alam et al., 2012). Statistical significance was determined using ANOVA. Internal standard curves were generated for each primer set by cloning the amplicon into a pGEM-Teasy vector (Promega) according to the manufacturer's instructions. Negative controls were included in all amplification reactions.

5.3.7 Fluorescent *In Situ* Hybridization

G. f. fuscipes specimens from the Seibersdorf lab colony were used for FISH. Teneral male and female flies were dissected in PBS 2-3 days after eclosion. Dissected tissues were dried on poly-L-lysine-coated glass slides (Sigma, UK) for 20 minutes at 65 °C and kept at 4 °C until further use. Tissue samples were fixed in freshly prepared 4 % paraformaldehyde solution for 30 minutes

at 4 °C. After fixation, slides were rinsed twice with PBS. To permeabilize the membranes, slides were coated at 60 °C with a few drops of 70 % acetic acid, rinsed with PBS after a 1-minute incubation, dehydrated through a graded ethanol series and left on the bench until completely dry. Then, slides were incubated in 0.2 % TritonX-100 in PBS for 2 minutes at room temperature, rinsed with PBS, dehydrated in ethanol, and air-dried. Pre-hybridization occurred for 15 minutes in pre-hybridization buffer containing 20 mM Tris-HCl pH 8.0, 0.9 N NaCl, 0.01 % sodium dodecyl sulphate and 30 % formamide. Hybridization was performed overnight at 35 °C by coating slides with 100 µl of pre-hybridization buffer supplemented with 1 µg ml^{-1} Cy3-labelled Spr403 probe targeting *Spiroplasma* (Matsuura et al, 2014). From this step on, all slides were kept in the dark. After hybridization, slides were rinsed in pre-hybridization buffer twice at 42 °C for 30 minutes each, followed by a quick rinse in PBS and in deionized water. The slides were air-dried, mounted with VECTASHIELD® mounting medium containing 1.5 µg ml^{-1} DAPI (Vector Laboratories, UK) and kept in the dark at 4 °C until observation on an Axio Observer.Z1 inverted microscope (Zeiss, Germany). Images were captured using an ApoTome.2 imaging system for optical sectioning, AxioCam MRm camera and Zen imaging software (Zeiss, Germany).

5.3.8 Whole genome shotgun sequencing and data analysis

Whole Genome Shotgun (WGS) libraries were prepared for five samples that were identified to contain novel taxa by 16S rRNA gene amplicon sequencing using the same gDNA. Fragment libraries were prepared with the Illumina TruSeq Nano DNA kit following the manufacturers' instructions. Samples were sequenced on two lanes of Illumina HiSeq with 250 bp paired-end reads. Raw sequencing reads were de-multiplexed and converted to FASTQ format with CASAVA version 1.8 (Illumina, 2011). Cutadapt version 1.2.1 (Martin, 2011) was used to trim Illumina adapter sequences from FASTQ files. Reads were trimmed if 3 bp or more of the 3' end of a read matched the adapter sequence. Sickle version 1.200 (Joshi and Fass, 2011) was used to trim reads based on quality: any reads with a window quality score of less than 20, or which were less than 10 bp long after trimming, were discarded.

Metagenomic reads were assembled with Discovar (Gmed5, Gpg3, and Gms8) and Abyss (Gmed4 and Gt6). Contigs shorter than 500 bp were removed and mapping with Bowtie2 was used to assess coverage. Taxonomy was assigned to contigs with BLAST and the GC content of contigs assessed with the Blobology package (Kumar et al., 2013). All non-bacterial contigs were filtered from the assembly and TAGC plots were drawn in R version 3.2.3 (R Core Team). The statistics presented in Table 5.4 correspond to contigs re-assembled from reads that mapped to contigs of target organisms. Contigs were filtered based on GC content, coverage and taxonomy, and reads were extracted using scripts implemented in Blobology. Extracted reads were re-assembled with SPAdes version 3.7.1 (Nurk et al., 2013) and mapped to contigs with Bowtie2. Assembly statistics were calculated with custom perl scripts and Qualimap version 2.2 (Okonechnikov et al., 2016). Assemblies were annotated with PROKKA version 1.5.2 (Seemann, 2014), and completeness was assessed following the method in (Rinke et al., 2013). Completeness values refer to the proportion of genes present in the assembly compared to the expected number in free-living bacteria.

5.4 Results

5.4.1 16S rRNA gene amplicon sequencing reveals novel interspecific diversity in natural populations of tsetse flies

Microbial community composition and diversity of thirty-two whole insects from four *Glossina* species (*G. medicorum, G. morsitans, G. p. gambiensis* and *G. tachinoides*) collected in Folonzo, Burkina Faso were investigated by 16S rRNA gene amplicon sequencing on the Illumina MiSeq platform. V4 region amplicons of the 16S rRNA gene were sequenced producing 5,761,899 reads after quality filtering. These reads were combined with a total of 8,300,515 quality-filtered reads generated from 124 whole guts of three tsetse fly species (*G. f. fuscipes, G. m. morsitans, G. pallidipes*) from a previous study by Aksoy *et al.* (2014), which used an identical technical approach for amplicon generation and sequencing. All sample details are listed in Table 5.1.

The primary nutritional endosymbiont of tsetse flies *Wigglesworthia* was the most abundant taxon in all samples, and constituted between 71 and 99 % of the total 16S rRNA copies in every individual. Variation in the relative abundance of *Wigglesworthia* was due to the heterogeneous distribution of secondary taxa, which varied in infection frequency and abundance between individuals in both an intra- and inter-specific fashion (Figure 5.1). Secondary taxa included the facultative symbionts *S. glossinidius* and *Wolbachia,* alongside several other genera including *Spiroplasma* and *Rickettsia*, which have not previously been reported in tsetse flies. The relative abundance of secondary taxa was highly variable (from < 0.01 to 28%) depending upon the genus of bacterium and the species of *Glossina* (Figure 5.1). This contributed to the variation in bacterial community composition between *Glossina* species. Clustering by species is illustrated in Figure 5.2, where Principal Component 1 and Principal Component 2 describe 58.53 % and 10.84 % of the variance respectively. Clustering can be partly attributed to the co-diversification of *Wigglesworthia*, which is the main component of the community, with its tsetse host (Chen et al., 1999). For this reason, outliers are conspicuous, as is

observed with the two individuals infected with *Spiroplasma* and *Rickettsia* at 13.15 % and 23.72 % relative abundance respectively (Figure 5.2).

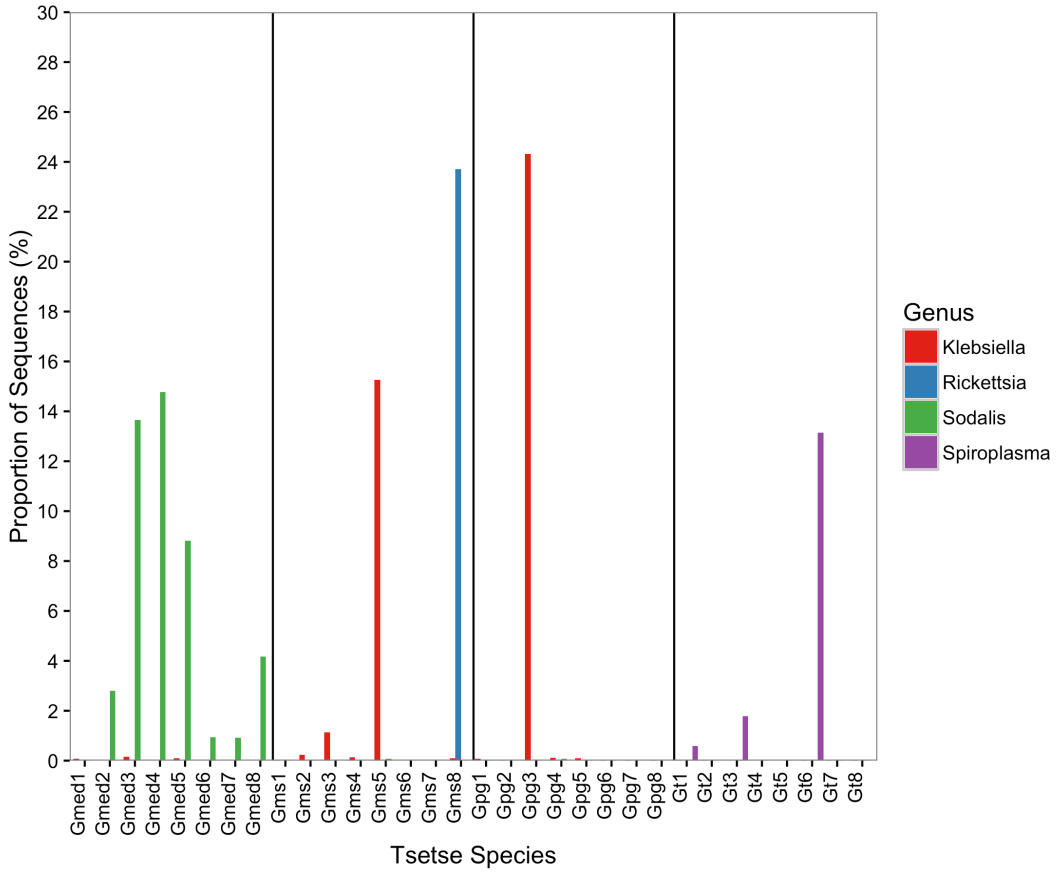


Figure 5.1 | Relative abundance of *Klebsiella*, *Rickettsia*, *Spiroplasma*, and *Sodalis* 16S rRNA copies in comparison to the total bacterial community in whole wild tsetse flies. Gmed: *G. medicorum*; Gms: *G. morsitans submorsitans*; Gpg: *G. p. gambiensis*; Gt: *G. tachinoides*.

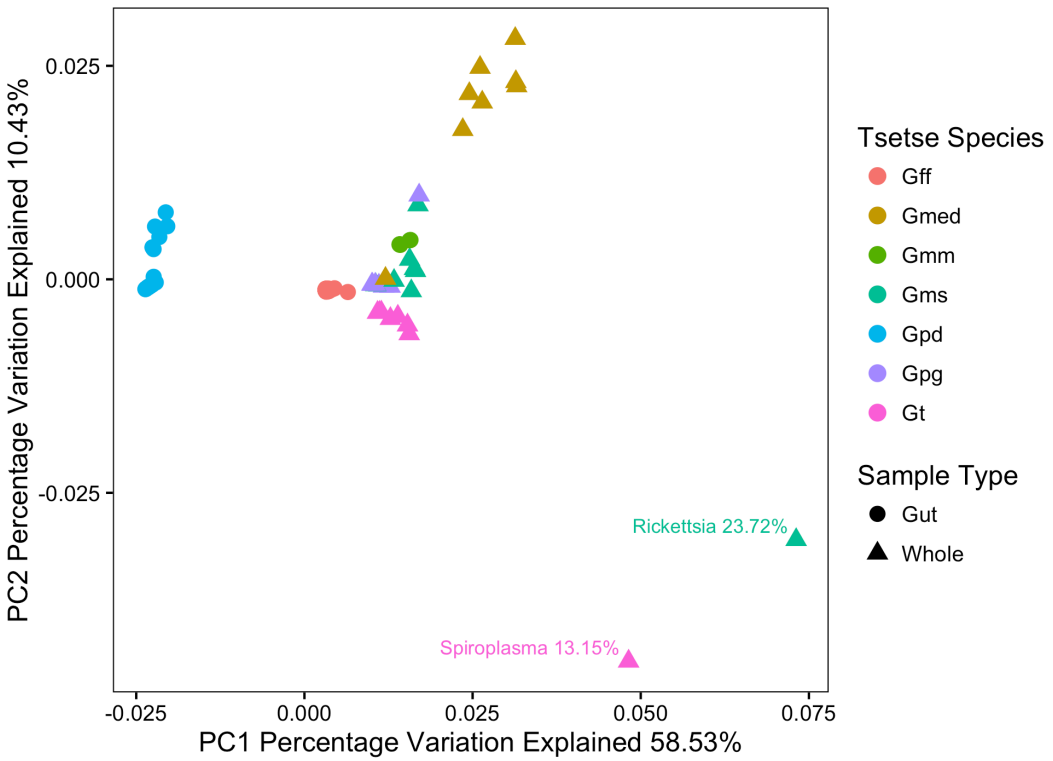


Figure 5.2 | Weighted Unifrac Principal Component Analysis of 16S rRNA gene amplicon data. Each data point represents an individual tsetse fly and is coloured according to *Glossina* species. All gut samples originated from the study by Aksoy *et al.* in 2014, and whole samples were collected in Burkina Faso. Gff: *G. fuscipes fuscipes*; Gmed: *G. medicorum*; Gms: *G. morsitans submorsitans*; Gpd: *G. pallidipes*; Gpg: *G. palpalis gambiensis*; Gt: *G. tachinoides*.

Sodalis was found at higher frequency and relative abundance in whole *G. medicorum* and *G. pallidipes* guts than in the other tsetse species studied (Figure 5.3). All individuals from other natural populations were either uninfected or infected with *Sodalis* at a relative abundance of 0.5 % or less. *Wolbachia* infections were found infrequently and at low abundance, regardless of whether whole flies or guts were examined: relative abundance did not exceed 0.1 % in any wild sample. Several other taxa previously associated with tsetse flies were detected including multiple members of the *Enterobacteriaceae*, such as *Klebsiella, Erwinia, Trabulsiella, Pantoea,* and *Serratia*. These infections occurred at low relative abundance, excluding those with *Klebsiella*, which was found in one *G. p. gambiensis* and one *G. m. submorsitans* whole fly at a relative abundance of 24.3 % and 15.3 % respectively. Amplicon profiling was also able to detect taxa that had not

previously been associated with the tsetse fly. Multiple wild individuals of *G. f. fuscipes* and *G. tachinoides*, which belong to the *palpalis* subgroup of the *Glossina* genus, were infected with *Spiroplasma.* Relative abundances were generally low (<1 %), but were found to be as high as 13.2 % in one *G. tachinoides* whole fly from Burkina Faso (Figure 5.1). Next-generation sequencing of the bacterial communities associated with whole insects revealed previously undiscovered diversity. Communities were characterized by infrequent but high abundance infections with several secondary taxa that have previously been shown to have a significant effect upon insect biology, including *Spiroplasma, Klebsiella* and *Rickettsia*.

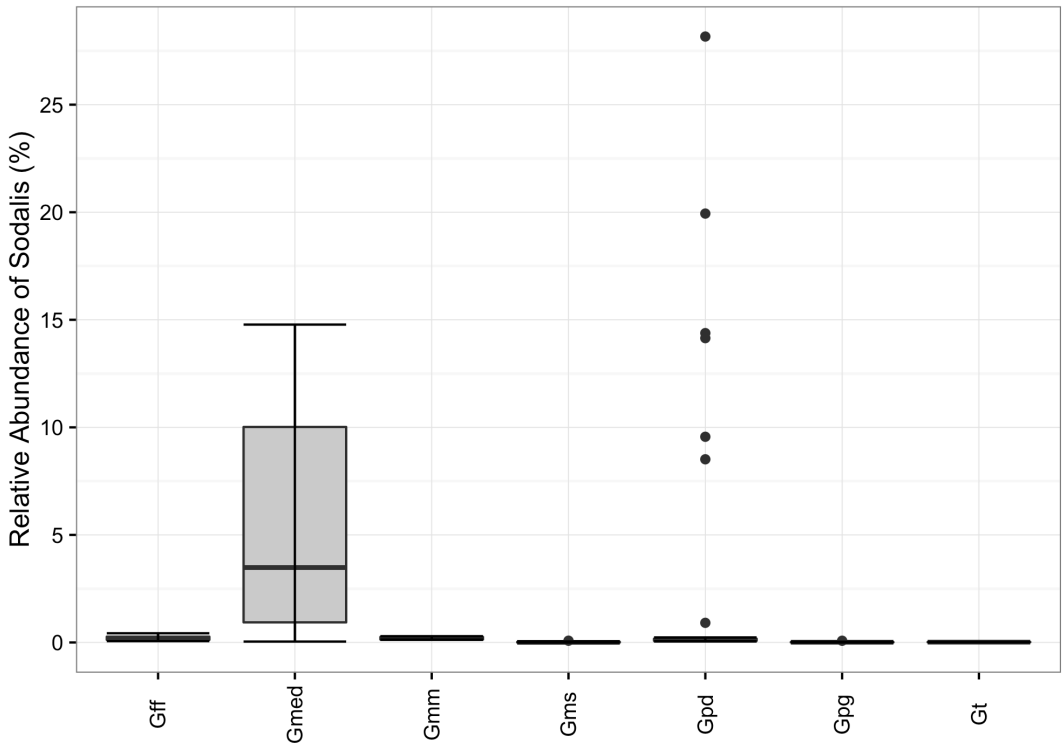


Figure 5.3 | *Sodalis* relative abundance in each tsetse species. Boxes denote the interquartile range, the line within the box in the median, and whiskers extend to the most extreme value within 1.5*interquartile range.

5.4.2 16S rRNA gene amplicon sequencing of laboratory reared tsetse flies

Microbial community composition in laboratory-reared tsetse flies was investigated by amplicon sequencing of the V3-V4 region of the 16S rRNA gene on the Illumina MiSeq platform. A total of 78 *G. f. fuscipes, G. m. morsitans* and *G. pallidipes* tissue samples from three developmental stages were sequenced, producing 2,445,369 reads after quality filtering. Similarly to wild populations, the three known taxa (*Wigglesworthia, Sodalis* and *Wolbachia*) were found in the laboratory flies. However, additional bacterial species were also detected, with members of *Flavobacterium*, *Propiniobacterium*, *Brevudimonas*, *Aeromonas*, and *Rhodospirillales* identified in *G. m. morsitans*, *G. f. fuscipes*, and *G. pallidipes*. Sequences related to *Acinetobacter*, and *Pantoea* were identified in *G. m. morsitans* and *G. pallidipes*. Additionally, sequences related to *Streptococcus* were found in *G. m. morsitans*, and *G. f. fuscipes*, while sequences related to *Shewanella*, and *Pedobacter* were found only in *G. m. morsitans*. Relative abundance was influenced by tissue sample type with gut tissues being enriched for *Wigglesworthia* and reproductive tissues for *Wolbachia* and *Sodalis*. We observed variation in the frequency and relative abundance of *Wolbachia* in lab populations. The mean relative abundance of *Wolbachia* was significantly higher in *G. m. morsitans* flies compared with those from the *G. f. fuscipes* or *G. pallidipes* populations (ANOVA, p = < 0.01) (Table 5.2). This was due to increased relative abundance of *Wolbachia* in reproductive tissues compared to larval or gut tissues within the *G. m. morsitans* population (ANOVA, p = < 0.01).

Table 5.2 | Mean relative abundance of *Wolbachia* and *Spiroplasma* in larval guts and the guts and reproductive tissues of teneral and 15-day-old adults from three species of tsetse fly reared in the laboratory (*G. f. fuscipes, G. m. morsitans,* and *G. pallidipes*).

Species (number of	*Wolbachia* mean relative	*Spiroplasma* mean relative
G. f. fuscipes (27)	0.03 ± 0.01	2.74 ± 1.05
G. m. morsitans (27)	1.65 ± 0.36	<0.01 ± <0.01
G. pallidipes (27)	0.23 ± 0.11	<0.01 ± <0.01

Spiroplasma relative abundance was significantly higher in *G. f. fuscipes* than *G. m. morsitans* and *G. pallidipes* flies (ANOVA, p = < 0.01) (Table 5.2). There was no significant difference in the mean relative abundance of *Spiroplasma* between sample types within the *G. f. fuscipes* population (ANOVA, p = 0.052) (Figure 5.4).

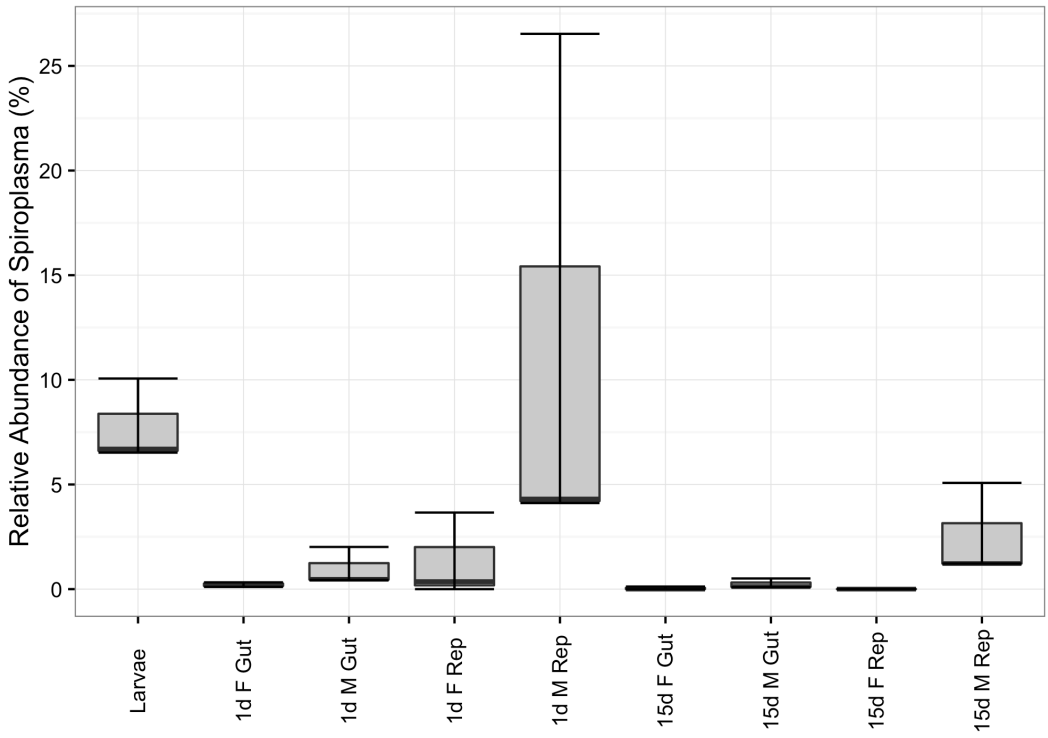


Figure 5.4 | *Spiroplasma* relative abundance in a laboratory colony of *G. f. fuscipes* flies of eight sample types: larvae, guts and reproductive organs from teneral and 15-day-old males and females. ANOVA revealed a non-significant difference between sample types (p = 0.052).

5.4.3 PCR screening of natural and laboratory tsetse populations

We used PCR-based screening methods to assay for the presence of four insect reproductive parasites: *Spiroplasma*, *Arsenophonous*, *Rickettsia*, and *Cardinium,* in four *Glossina* species from the lab and wild: *G. m. morsitans* (n=19), *G. f. fuscipes* (n=174), *G. pallidipes* (n=20), and *G. p. gambiensis* (n=19). All host DNA samples were positive for PCR amplification using the 12S rRNA gene arthropod universal primers (data not shown), indicating satisfactory DNA quality. Of the four examined *Glossina* species, *Spiroplasma* infections were only found in *G. f. fuscipes*, while none of the four tsetse

species were infected with *Arsenophonus*, *Rickettsia* or *Cardinium*. To examine the distribution of *Spiroplasma*, six additional *Glossina* species *G. austeni* (n=32), *G. brevipalpis* (n=21), *G. m. centralis* (n=1), *G. m. submorsitans* (n=8), *G. p. palpalis* (n=16), and *G. tachinoides* (n=15) were PCR-screened for *Spiroplasma* infection. Only *G. tachinoides* and *G. p. palpalis* were positive for *Spiroplasma*, and showed an infection rate of 26.7 % and 12.5 % respectively (Table 5.3). The PCR screening for *Spiroplasma* infection was further extended to 327 historical and contemporary samples from wild and lab colonies representing 10 species of tsetse fly (Table 5.3). Only members of the *palpalis* subgroup were infected with *Spiroplasma*, including *G. fuscipes fuscipes, G. palpalis palpalis* and *G. tachinoides*, with a prevalence ranging from 6 % to 80 %. Notably, the prevalence was higher in laboratory colonies than natural populations, and some populations demonstrated a disparity in infection between sexes (Table 5.3).

5.4.4 Genotyping of *Spiroplasma* strains

Spiroplasma strains from *G. f. fuscipes* flies of both sexes from the Seibersdorf and Bratislava lab colonies, a natural population from Uganda (Lukoma-Buvuma islands, female only) and from one natural population of *G. tachinoides* flies from Burkina Faso (female only) were genotyped by MLST analysis. Both laboratory and field samples of *G. f. fuscipes* harboured *Spiroplasma* strains with identical sequences for all loci studied (Table 7.5). Interestingly, the *Spiroplasma* strain present in *G. tachinoides* is distinct from the *G. f. fuscipes Spiroplasma* strain with sequence polymorphisms detected in all loci examined. Eight polymorphisms were observed in *fruR*, seven in the region 16S *rRNA*-23S *rRNA*-5S *rRNA*, four in 16S rRNA, three in *dnaA*, two in *ftsZ*, and one in *rpoB* and *parE*. Both strains belong to the Citri clade (Figure 5.5). This topology was confirmed by all examined genetic markers (Figure 7.1 - 7.6). Unfortunately, a lack of *Spiroplasma* gDNA prevented further characterization of the strain present in *G. p. palpalis*.

Table 5.3 | *Spiroplasma* prevalence in ten *Glossina* species assessed by PCR in natural and laboratory populations.

Species	Origin	Collection date	Location (area, population)	Tissue	Number of samples	Spiroplasma infection frequency
G. austeni	Field	1995	Tanzania (Zanzibar)	Whole	10	0
	Field	1996	Tanzania (Jozani)	Whole	10	0
	Field	1999	South Africa (Zululand)	Whole	10	0
	Field	Unknown	Coastal Tanzania (Muhoro)	Whole	2	0
G. brevipalpis	Lab	1995	Seibersdorf Lab Colony	Whole	16	0
	Lab	Unknown	Coastal Tanzania (Pangani)	Whole	5	0
G. f. fuscipes	Field	1994	Uganda (Buvuma Island, GFTF2)	Whole	17	0
	Field	1994	Uganda (Buvuma Island, GFKF2)	Whole	5	0
	Field	1994	Uganda (Buvuma Island, GFFBUV2)	Whole	9	0
	Field	1994	Uganda (Buvuma Island, GFFTOR2)[3]	Whole	15	6.7
	Lab	1995	Seibersdorf Lab Colony[1]	Whole	36	33.4
	Lab	2013	Bratislava Lab Colony[2]	Whole	40	80
	Field	2014	Uganda (Lukoma-Buvuma Islands, 350)[2]	Whole	52	5.8
G. m. centralis	Lab	2008	Yale Lab Colony	Whole	1	0
G. m. morsitans	Lab	2008	KARI-TRC Lab Colony	Whole	15	0
	Lab	2010	Antwerp Lab Colony	Whole	4	0
G. m. submorsitans	Field	2010	Burkina Faso (Folonzo)	Whole	8	0
G. pallidipes	Lab	1999	Seibersdorf Lab Colony	Whole	2	0
	Lab	2008	Seibersdorf Lab Colony	Whole	13	0
	Lab	Unknown	Uganda-UGA/IAEA	Whole	5	0
G. p. gambiensis	Lab	1995	CIRDES Lab Colony	Whole	9	0
	Lab	2005	CIRDES Lab Colony	Whole	10	0
G. p. palpalis	Lab	1995	Seibersdorf Lab Colony[3]	Whole	16	12.5
G. tachinoides	Lab	1995	Seibersdorf Lab Colony[3]	Whole	7	14.3
	Field	2010	Burkina Faso (Folonzo)[2]	Whole	8	37.5

5.4.5 *Spiroplasma* density across developmental stages

qPCR was used to assess the density of *Spiroplasma* infection in larval guts as well as in guts and gonads of teneral (newly-eclosed and unfed) and 15-day-old male and female adult flies. *Spiroplasma* infection levels were significantly higher in larval guts compared to the guts of teneral or 15-day-old adults (Figure 5.6A). There was no significant difference in the infection levels

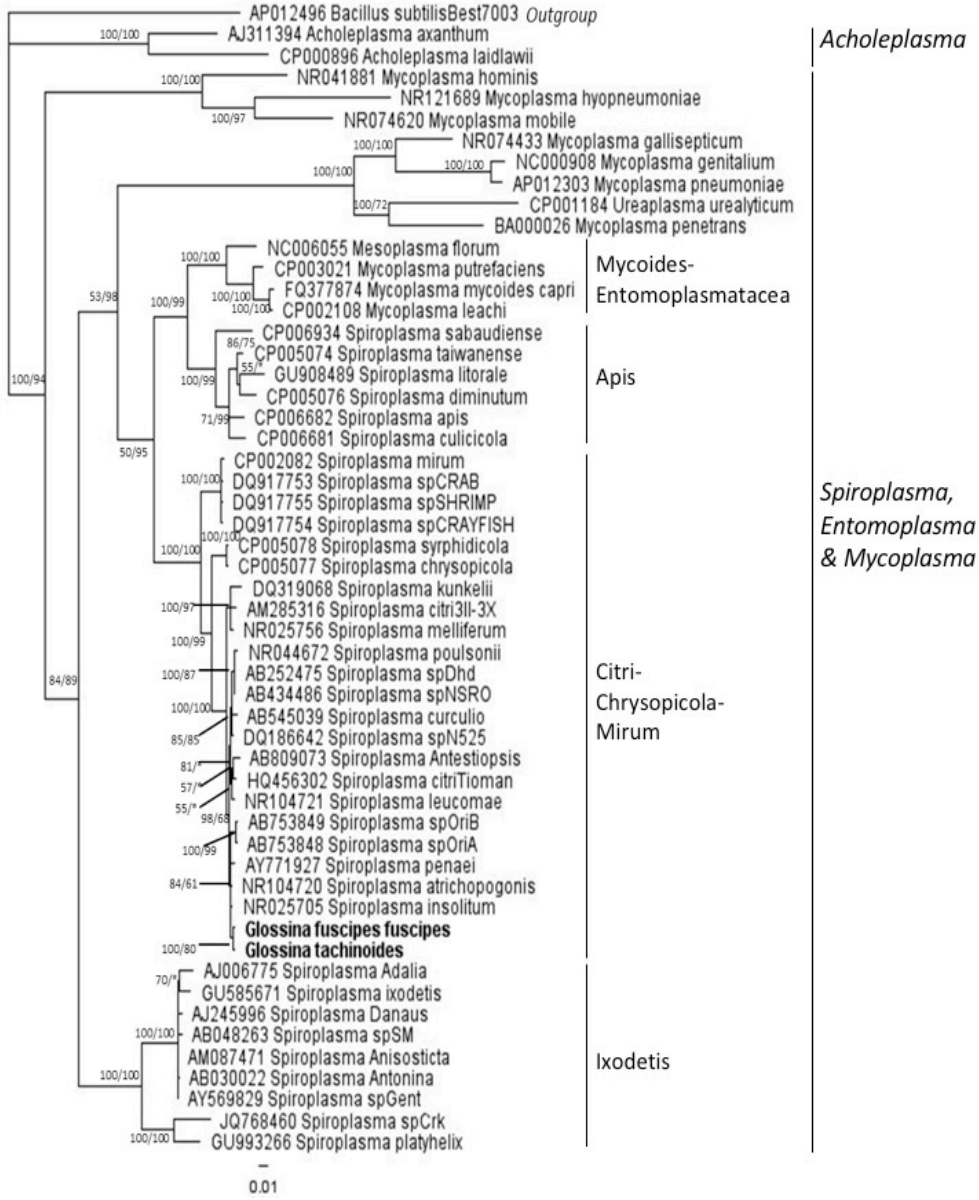


Figure 5.5 | Bayesian inference phylogeny of the 16S rRNA gene of the *Glossina fuscipes fuscipes* and *G. tachinoides Spiroplasma* strains. Maximum Likelihood (ML) topology was similar. Bayesian posterior probabilities and ML bootstrap values based on 1000 replicates are given at each node (only values >50% are indicated), respectively. Asterisks indicate support values lower than 50%.

between testes of teneral and 15-day-old adults (Figure 7.7). However, there was a significant difference in *Spiroplasma* infection level between testes and ovaries from teneral flies (Figure 5.6B).

This corresponded with results from 16S rRNA gene amplicon profiling of the same tissues (Figure 5.4). *Spiroplasma* density was also examined in a mass-reared colony where lethality was high and the colony was on the verge of collapse. We examined live and dead insects, and results indicated that in males *Spiroplasma* density was similar, whereas in females, density was higher in live insects than in those that had recently died (Figure 5.7A and Figure 5.7B). When we examined females carrying a larva, we found that the live females with a larva had a higher titre of *Spiroplasma* than gravid females that died prematurely (Figure 5.7C).

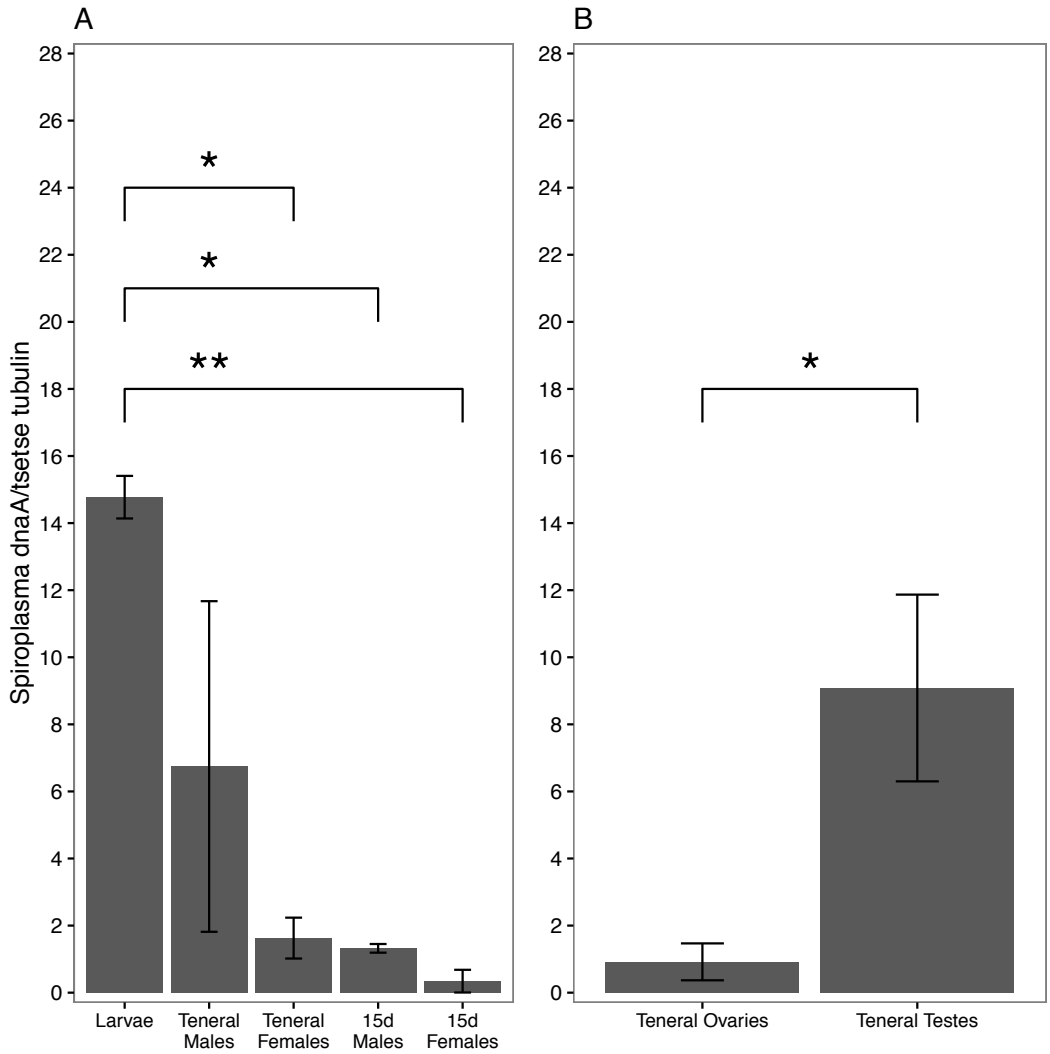


Figure 5.6 | *Spiroplasma dnaA* copy number normalized to the *Glossina β-tubulin* gene in **A** the guts and **B** the teneral reproductive tissues of *G. f. fuscipes* laboratory flies. Three biological replicates of five pooled individuals were assessed for each life stage. (*p<0.05, **p<0.01, ANOVA).

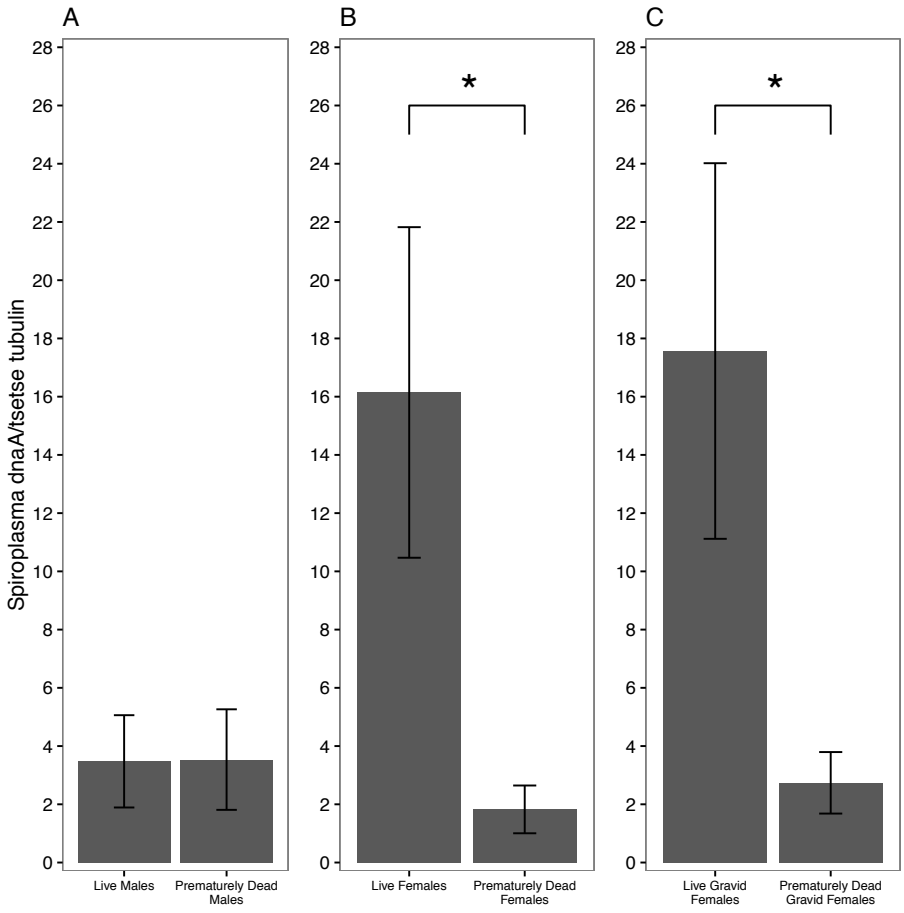


Figure 5.7 | *Spiroplasma dnaA* copy number normalized to the *Glossina β-tubulin* gene. *Spiroplasma* titre in living and dead whole *G. f. fuscipes* **A** males **B** females **C** gravid females from a mass-rearing colony in Ethiopia. (*p<0.05, ANOVA).

5.4.6 Fluorescent In Situ Hybridization of *Spiroplasma*

Dissected ovaries and testes of teneral adults from a *G. f. fuscipes* laboratory colony were subjected to fluorescence *in situ* hybridization using *Spiroplasma* specific probes. *Spiroplasma* was detected in testes at high densities (Figure 5.8A), whereas localization of *Spiroplasma* was sparse and sporadic in ovaries (Figure 5.8B).

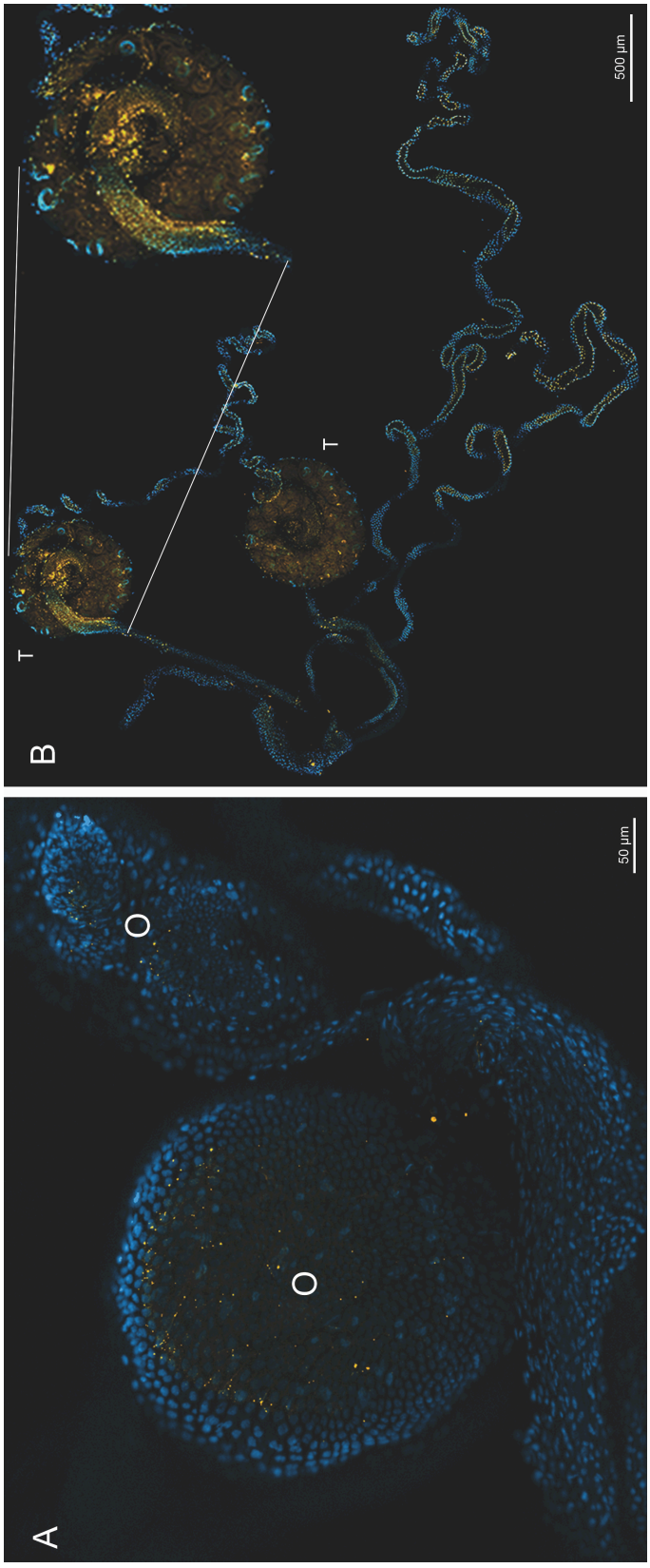


Figure 5.8 | Localization of *Spiroplasma* in the reproductive tissues of female and male *G. f. fuscipes* adults by FISH. Insect nuclear DNA is stained in blue and *Spiroplasma* in yellow. **A** FISH of *G. f. fuscipes* ovaries (O) **B** FISH of *G. f. fuscipes* testes (T).

5.4.7 Identification of novel taxa by whole genome sequencing

Samples for shotgun sequencing were selected based on the relative abundance of novel taxa identified by 16S rRNA gene sequencing (Figure 5.9). These were *Sodalis* from two *G. medicorum* individuals (Figure 5.10, Figure 5.11), *Rickettsia* from one *G. m. submorsitans* individual (Figure 5.12), *Klebsiella* from one *G. p. gambiensis* individual (Figure 5.13), and *Spiroplasma* from one *G. tachinoides* individual (Figure 5.14). Assemblies of varying quality were generated from metagenomic sequencing of all targeted taxa (Table 5.4). Validation with WGS indicates the effectiveness of 16S rRNA gene profiling as a method for detecting novel members of the microbiota.

There was broad variation in the size, GC content, and quality of assemblies produced from novel taxa (Table 5.4). Despite its high relative abundance according to 16S rRNA gene profiling, the genome assembly for the *Klebsiella* isolate from *Gpg3* was of significantly lower quality than those from other taxa. Assembly statistics indicate that only 15.1 % of the predicted *Klebsiella* genome is represented in the current assembly (0.9 Mb), and that it is highly fragmented (505 contigs, N50 1648). All remaining taxa were over 90 % complete in comparison to free-living bacteria. This value may be underestimated for *Rickettia* and *Spiroplasma*, which tend to be intracellular bacteria with reduced genomes. For example, complete genomes from *Wigglesworthia glossinidia*, which are highly reduced due to the evolutionary dynamics of intracellular symbionts, score only 92 % in comparison to free-living bacteria. These values may therefore need to be adjusted to compensate for ecology and the age of associations. The observed variation in size and GC content between assemblies corresponds with the putative lifestyles of each isolate, which are predicted from the roles of similar taxa in other arthropods.

Table 5.4 | Assembly statistics for the draft genomes of novel taxa or strains of bacteria isolated from whole genome shotgun-sequencing data from whole individual tsetse flies.

Sample Name	**Gmed4**	**Gmed5**	**Gms8**	**Gpg3**	**Gt6**
***Glossina* Species**	*G. medicorum*	*G. medicorum*	*G. m. submorsitans*	*G. p. gambiensis*	*G. tachinoides*
Target Organism	*Sodalis*	*Sodalis*	*Rickettsia*	*Klebsiella*	*Spiroplasma*
Number of Reads	1,051,788	803,368	793,004	92,986	253,072
Assembly Size (MB)	3.95	3.88	1.45	0.9	1.23
Number of Contigs	1091	1092	172	505	219
N50	8383	7998	14117	1648	13762
GC %	50.58	51.21	32.41	55.24	29.9
Mean Coverage	31.9	25.63	58.78	11.66	13.07
Number of RNA Genes	60	55	35	21	34
Number of CDS	5149	5204	1538	959	2130
Completeness (%)	97.84	97.12	92.09	15.11	91.37

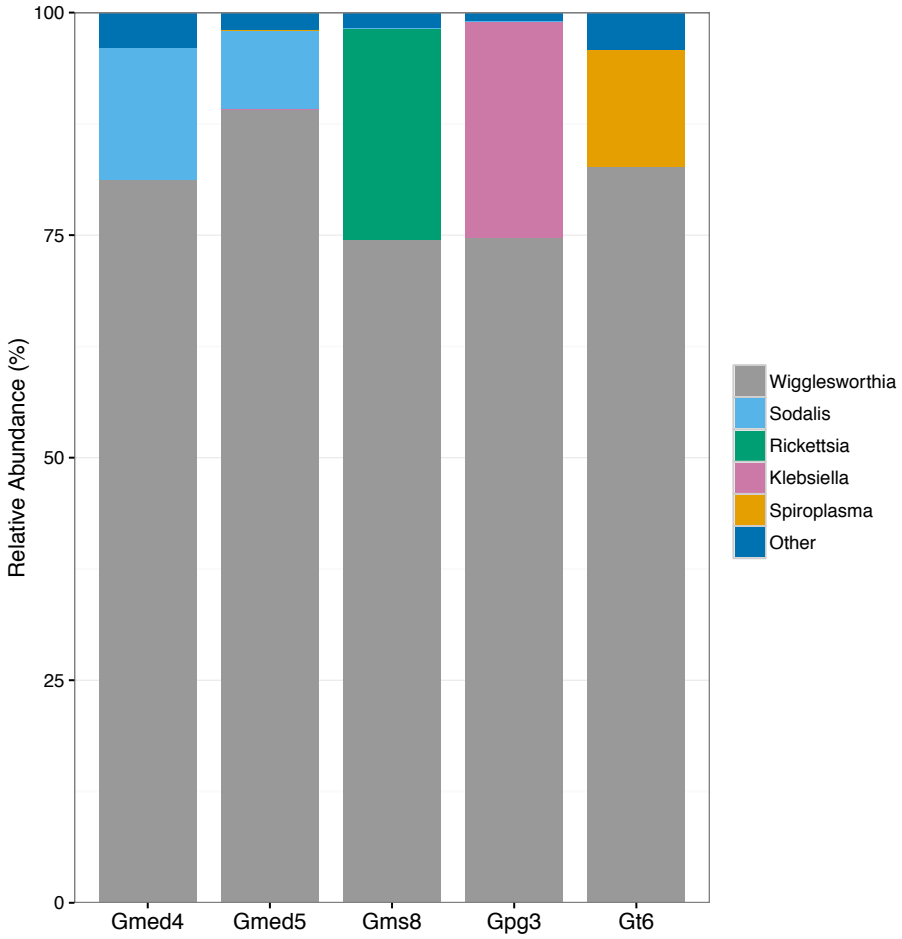


Figure 5.9 | The relative abundance of 16S rRNA gene copies assessed by amplicon profiling in samples identified for tsetse fly whole genome shotgun sequencing to obtain draft genomes for bacterial taxa of interest. Details of sample type and taxa of interest are noted in Table 5.4.

5.4.8 Whole genome sequencing of *Wigglesworthia* and *Wolbachia*

In addition to novel taxa, *Wigglesworthia* was detected in all shotgun libraries (Figure 5.10-Figure 5.14). Despite the very low relative abundance of *Wolbachia* in 16S rRNA gene profiling experiments of natural populations (data not shown), it was also found in varying abundance in all individuals. Individual Gms8 infected with *Rickettsia* hosted the highest number of *Wolbachia* contigs (Figure 5.12). This may have been caused by chromosomal insertions of *Wolbachia* DNA into the host genome, as has previously been documented in *Morsitans* group flies (Brelsfoard et al., 2014), misassignment of *Rickettsia* contigs to *Wolbachia* due to their close phylogenetic relationship, or to a double infection with both *Rickettsia* and *Wolbachia*.

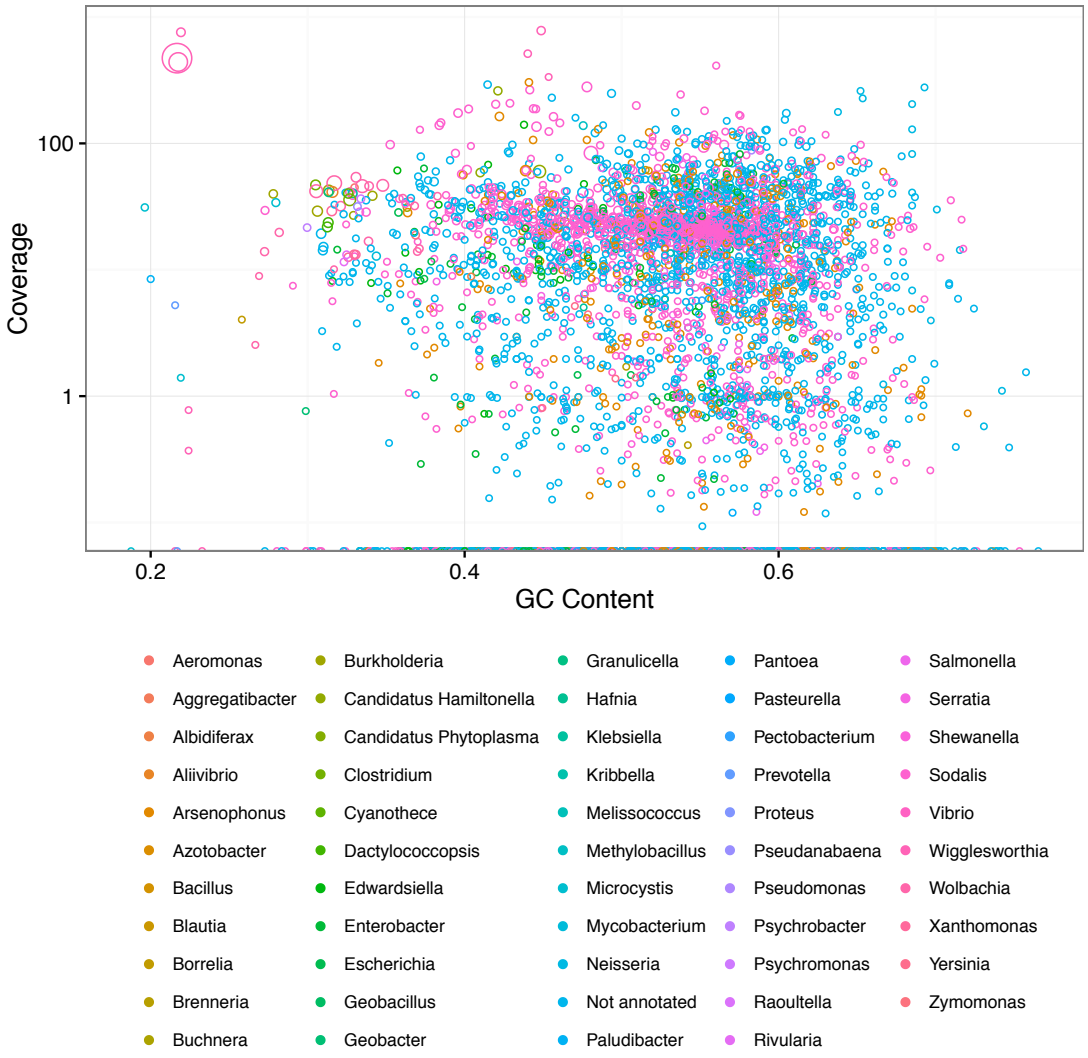


Figure 5.10 | Taxon-Annotated GC Coverage (TAGC) plot of all bacterial contigs associated with sample Gmed4. The size of data points is proportional to the length of the contig. The relative abundance of *Sodalis* estimated by 16S rRNA gene profiling was 14.8 %.

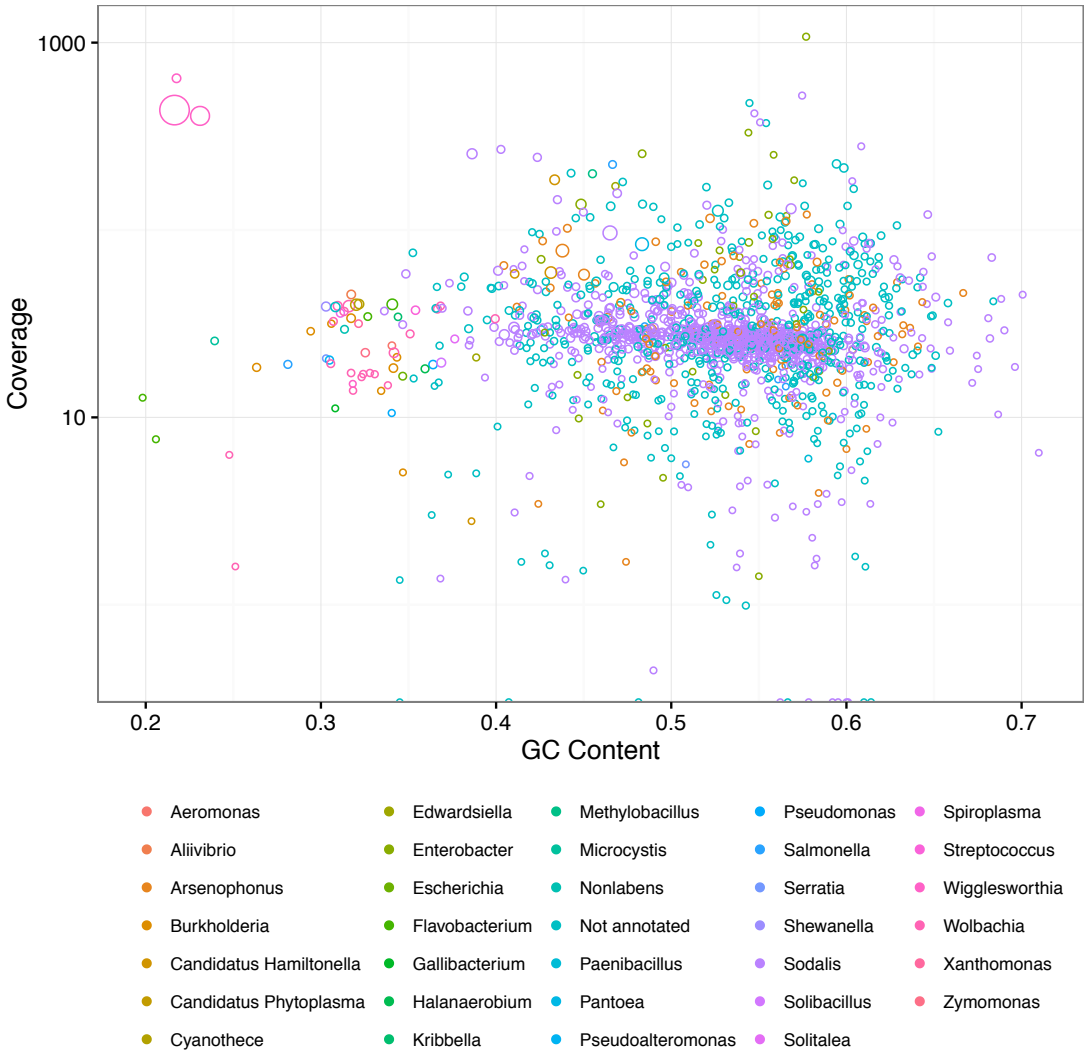


Figure 5.11 | TAGC plot of all bacterial contigs associated with sample Gmed5. The size of data points is proportional to the length of the contig. The relative abundance of *Sodalis* estimated by 16S rRNA gene profiling was 8.8 %.

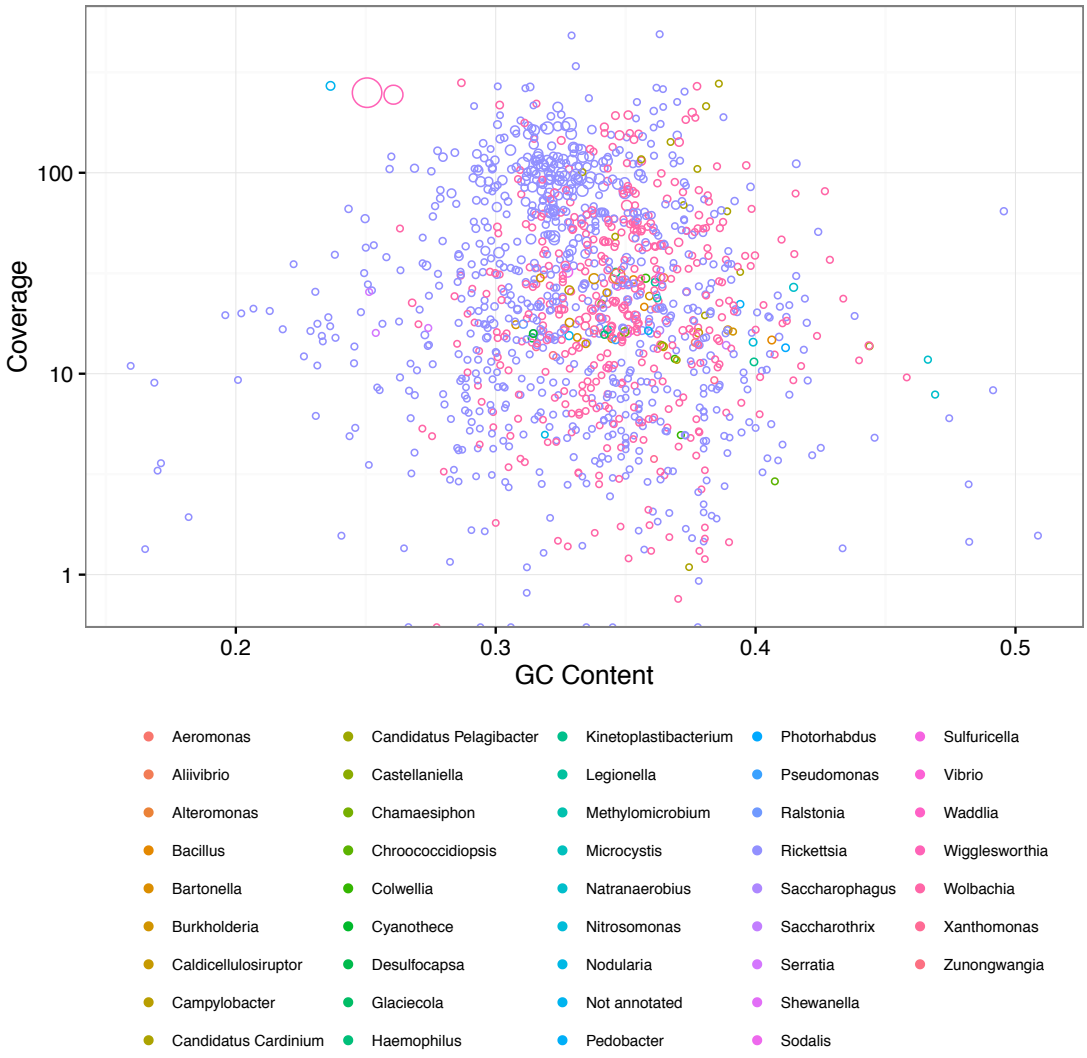


Figure 5.12 | TAGC plot of all bacterial contigs associated with sample Gms8. The size of data points is proportional to the length of the contig. The relative abundance of *Rickettsia* estimated by 16S rRNA gene profiling was 23.7 %.

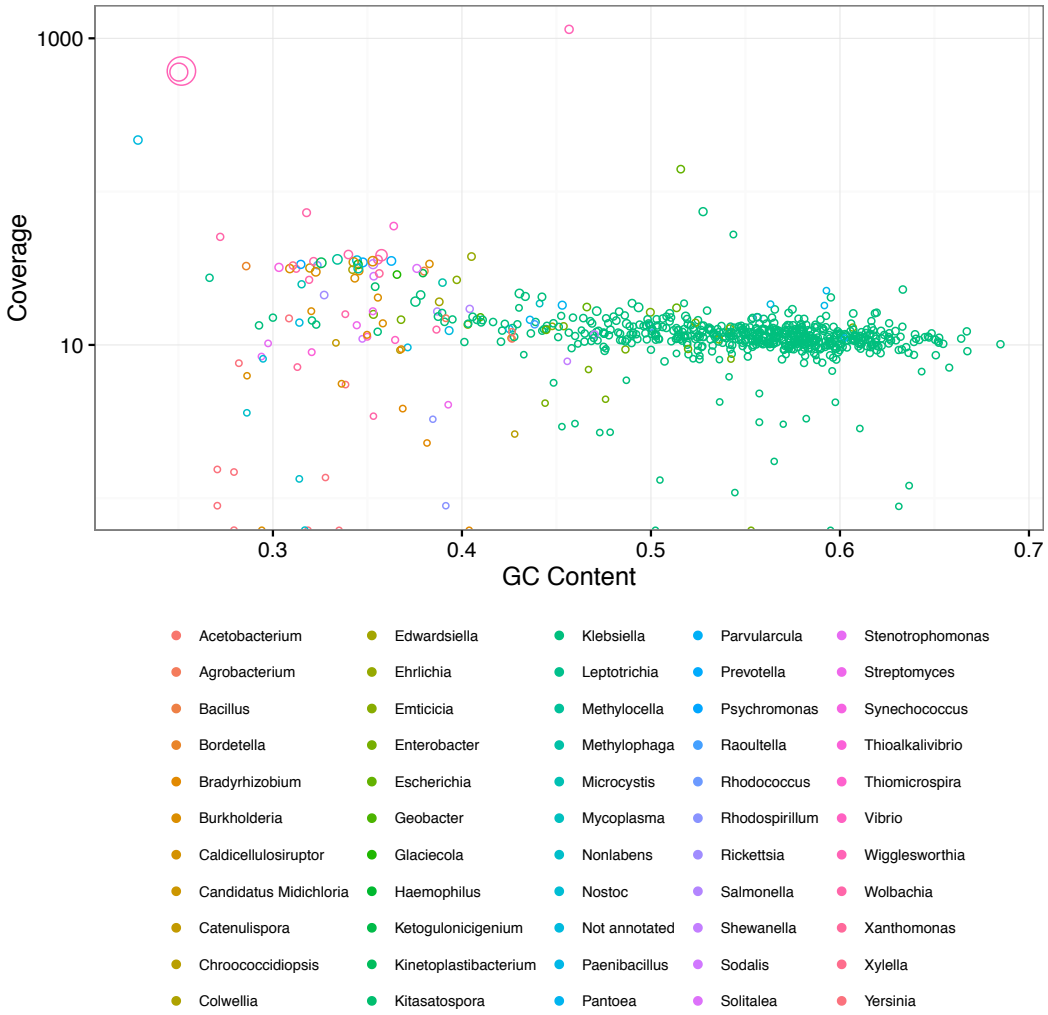


Figure 5.13 | TAGC plot of all bacterial contigs associated with sample Gpg3. The size of data points is proportional to the length of the contig. The relative abundance of *Klebsiella* estimated by 16S rRNA gene profiling was 24.3 %.

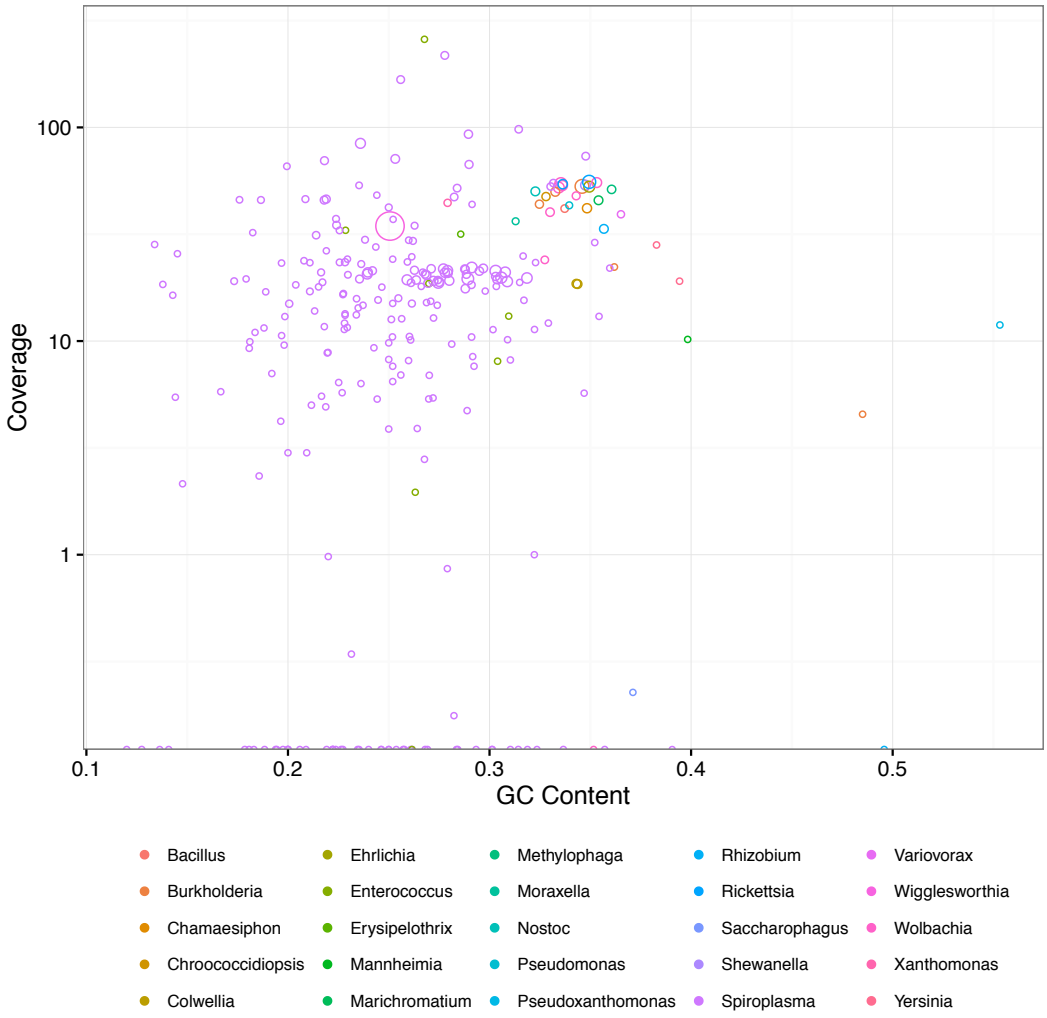


Figure 5.14 | TAGC plot of all bacterial contigs associated with sample Gt6. The size of data points is proportional to the length of the contig. The relative abundance of *Spiroplasma* estimated by 16S rRNA gene profiling was 13.2 %.

5.5 Discussion

5.5.1 Bacterial diversity in tsetse flies is higher than previously reported

The bacterial communities associated with tsetse flies are more complex than previously reported, thus challenging the *Wigglesworthia-Sodalis-Wolbachia* dogma (Wamwiri et al., 2013; Wang et al., 2013; Geiger et al., 2015). Using 16S rRNA gene amplicon sequencing several additional bacterial genera with broad phylogenetic origins were discovered to be associated with tsetse flies, including *Klebsiella, Rickettsia* and *Spiroplasma*. The prevalence and infection levels observed in some tsetse species, particularly those of *Spiroplasma*, were similar to those seen in *Sodalis*, suggesting that they may play an important role in the biology and ecology of tsetse flies.

Previous studies have shown that the microbiota is characterized by the presence of *Wigglesworthia*, *Sodalis* and *Wolbachia*. All three symbionts are maternally transmitted, and colonize during the early juvenile stages: *Wigglesworthia* and *Sodalis* through milk gland secretions as larvae, and *Wolbachia* through the germ line during embryogenesis (Balmand et al., 2013; Wang et al., 2013). As larvae are intrauterine, the only bacteria that they encounter prior to pupation originate from within the adult female tsetse fly. Due to the obligate requirement of *Wigglesworthia*, there is high fidelity in vertical transmission from mother to offspring (Attardo et al., 2008). This makes it difficult for other bacteria to invade, as microbes occupy many of the available niches within the host from the early stages of development. Conversely, this also means that the tsetse immune system has evolved to accommodate bacteria (Weiss et al., 2011), which could facilitate colonization by environmental microbes able to exploit deficits in the immune system. Due to the unique biology of tsetse flies, there is only a short time window for colonization between larval deposition and pupation in the soil. In addition, the colonizers would have to survive metamorphosis in order to persist. Until recently, there was the notion that tsetse flies feed exclusively on blood, which is mostly sterile and therefore should not serve as a source of microbes.

However, there is now evidence that *G. p. gambiensis* flies deprived of a blood meal can feed on water or sugar water, and that sugar residues are detectable in wild-caught flies (Solano et al., 2015). Therefore, it is possible that these previously unrecognized feeding habits could be a source of environmental microbes, and could be the origin of the low-frequency high-abundance infections observed in multiple individuals in this study.

Host genotype can play an important role in the structure of the associated symbiotic community. Tsetse species-specific trends were noticeable for the different bacterial genera identified in the present study. *Spiroplasma* was detected in members of the *palpalis* sub-group (*G. f. fuscipes, G. p. palpalis* and *G. tachinoides*), whereas *Sodalis* was more prevalent in *G. medicorum* (*fusca* group). Previous studies have also shown that *Sodalis* infection is more prevalent in *G. brevipalpis* (*fusca* group) than in *G. m. morsitans* and *G. pallidipes* (both *morsitans* group) (Dennis et al., 2014). The relationship of *Spiroplasma* with *palpalis* seems to be more exclusive than that of *Sodalis*, since the latter has previously been identified in individuals belonging to all tsetse sub-groups (Geiger et al., 2005; Aksoy et al., 2014; Dennis et al., 2014).

A key approach to detecting invasive taxa is to sample whole insects rather than individual tissues such as the gut, where *Wigglesworthia* is dominant and obscures the detection of lower-abundance taxa. As there is variation between sub-groups, species, and even individuals within the same species, host genotype is an important factor. There may also be spatial patterns in bacterial infection, so using a wide geographic range could help to uncover bacterial diversity. Greater sampling effort and more sensitive detection techniques will aid in the search for complex tsetse diversity, as primary and secondary symbiont titres are already very high, so uncovering the rarer and low-abundance taxa is difficult. For example, *Rickettsia* was discovered at high abundance in just one individual, despite the profiling of hundreds of insects by amplicon and PCR profiling during this study and that by Aksoy *et al.* (2014). Many of the taxa found in tsetse flies belong to the family *Enterobacteriaceae*, for which it is notoriously difficult to construct a

phylogeny using the 16S rRNA gene (Moran et al., 2008; Naum et al., 2008; Husník et al., 2011). While *Serratia* has been isolated from tsetse flies and identified by sequencing of the full-length 16S rRNA gene (Geiger et al., 2010; Lindh and Lehane, 2011), the V4 region does not provide enough phylogenetic information to reliably distinguish the strain(s) found in tsetse flies from *Sodalis* (data not shown). For example, Aksoy et al. found a co-association between *Sodalis* and *Serratia* at very low relative abundance (Aksoy et al., 2014). Lack of a robust phylogeny combined with the short length of the region used may have caused the misclassification of low abundance *Sodalis* OTUs to *Serratia.*

5.5.2 *Spiroplasma* as a new symbiont of tsetse flies

Spiroplasma was associated with both natural and laboratory populations of tsetse from species belonging to the *palpalis* sub-group (*G. f. fuscipes, G. p. palpalis* and *G. tachinoides*). PCR-screening experiments indicated that *Spiroplasma* infection was more prevalent in lab colonies and that both males and females harboured *Spiroplasma*, whereas in natural populations prevalence was lower and only females were infected. The lack of infection in wild individuals may be due to insufficient sampling effort or a lack of sensitivity in the detection methods. It has been reported, for example, that some symbionts may be present in such low abundances that they are undetectable by conventional PCR screens (Schneider et al., 2014). MLST indicated that the *Spiroplasma* strain found in wild *G. f. fuscipes* from Uganda was identical, based on the loci examined, to that in the colonized flies (originating from the Central African Republic). This suggests that the intitial *Spiroplasma* infection of *G. f. fuscipes* may have occurred prior to the splitting of these tsetse sub-groups, or that horizontal transmission may occur between tsetse sub-groups. Due to artificially imposed limits on population size and the close proximity in which tsetse flies are reared in the lab, transmission of secondary symbionts is hypothesized to be higher, which leads to higher prevalence of infection. Although there have been no direct studies on the relative transmission rate of tsetse symbionts in the laboratory and field, paternal transmission during mating can occur for the secondary symbiont *Sodalis* (De Vooght et al., 2015). While this study only detected

Spiroplasma infection in *palpalis* group flies, screening more specimens from the *morsitans* and *fusca* groups should provide more detailed information on the dynamics and spread of *Spiroplasma* infection in natural populations. In addition, its infection levels should be compared with those of the other tsetse symbionts.

Another potential explanation for the absence of *Spiroplasma* in the *morsitans* and *fusca* groups is their frequent infection with *Wolbachia* (Doudoumis et al., 2012). An existing *Wolbachia* infection may have led to the development of competitive exclusion with *Spiroplasma*, though it is not yet clear whether they share an ecological niche within the host, and whether co-occurrence could create evolutionary pressure strong enough to drive competitive exclusion (Charlat et al., 2006). In *D. melanogaster*, coinfections between *Wolbachia* and *Spiroplasma* were asymmetrical: *Spiroplasma* negatively affected the titre of *Wolbachia*, whereas *Wolbachia* density did not affect *Spiroplasma* titer (Goto et al., 2006). Similarly to *Spiroplasma* in *G. f. fuscipes*, tissue tropism was observed in *D. melanogaster* infected with *Spiroplasma*, where density was highest in the ovaries (Goto et al., 2006). Competitive inter- and intraspecific microbial interactions have also been observed in mosquitoes with the mutual exclusion of *Asaia* and *Wolbachia* in the reproductive organs, while the native gut microbiota seems to prevent the vertical transmission of *Wolbachia* in *Anopheles* mosquitoes (Hughes et al., 2014; Rossi et al., 2015). *G. f. fuscipes* are infected with *Wolbachia*, though prevalence in natural populations is very heterogeneous, with an average infection rate of 44.3% (Alam et al., 2012). *Wolbachia* titre is extremely low in *G. f. fuscipes* insects; possibly at the limits of detection, which may explain why other screening exercises have failed to detect this symbiont in natural *G. f. fuscipes* populations (Alam et al., 2012; Doudoumis et al., 2012). *Spiroplasma*, on the other hand, is found at much lower frequency in natural populations, but is at higher density per individual when compared with *Wolbachia*. A similar example has been observed between two primary endosymbionts in sharpshooters (*Cicadellidae*), where *Sulcia mulleri* and *Baumannia cicadellinicola*, have evolved complementary metabolic functions (Moran et

al., 2008) in *D. melanogaster*, as mentioned earlier (Goto et al., 2006), and in bees (Engel et al., 2012).

MLST analysis indicated that the *Spiroplasma* strains detected in *G. f. fuscipes* and *G. tachinoides* populations, albeit different, both belong to the *citri* clade. Prominent examples of taxa from this clade include *S. kunkelii*, *S. phoeniceum*, and *S. citri*, all of which are plant pathogens (Saglio et al., 1973; Saillard et al., 1987; Whitcomb et al., 1986). *S. poulsonii,* which has been shown to have a protective effect against parasitic wasps in *D. melanogaster,* is also a member of this clade (Paredes et al., 2015).

5.5.3 *Spiroplasma* infection dynamics in lab-reared *G. f. fuscipes*

Spiroplasma density was investigated by qPCR analysis of gut and reproductive tissues from a lab colony of *G. f. fuscipes* flies at multiple life stages. When examining gut tissues, *Spiroplasma* titre was highest in larvae, and gradually decreased in both males and females over the course of adulthood. High larval titre indicates vertical transmission from mother to offspring, possibly via the milk gland; a mechanism already exploited by *Wigglesworthia* and *Sodalis* (Attardo et al., 2008). High larval density is an abnormal trait in the context of other insect-associated Spiroplasmas. Multiple strains of *Spiroplasma* infect a number of species of *Drosophila* and are able to induce a variety of phenotypes in their insect host ranging from parasitic reproductive manipulators to protective symbionts (Anbutsu and Fukatsu, 2011; Paredes et al., 2015). However, density always peaks after metamorphosis (Goto et al., 2006). In *D. hydei* and *D. melanogaster, Spiroplasma* titre steadily increases during larval and adult development with no differentiation between males and females (Goto et al., 2006). Interestingly, *Drosophila* male killing *Spiroplasma* strains exhibit a very high titre in the haemolymph (Anbutsu and Fukatsu, 2011), a pattern not observed in the *G. f. fuscipes Spiroplasma* strain discovered here (data not shown). In addition, *Spiroplasma* titre in *G. f. fuscipes* is much lower than that described for *Drosophila* male killing strains (Anbutsu and Fukatsu, 2003; 2011). *Wolbachia* is the only other maternally inherited endosymbiont found in

Drosophila, and is also found in tsetse flies (Doudoumis et al., 2012). *Wolbachia* confers density-dependent protection against insect viruses at different developmental stages in several *Drosophila* species (Teixeira et al., 2008; Hedges et al., 2012; Martinez et al., 2014; Stevanovic et al., 2015). Given the ecological similarities between *Wolbachia* and *Spiroplasma,* it is possible that high *Spiroplasma* density may also play a role in larval fitness. This warrants further study, as protection against viral or bacterial pathogens during intrauterine larval development would consistute a rare phenotype for a bacterial endosymbiont.

Gut infection was maintained into adulthood, particularly in males, where *Spiroplasma* load was significantly higher than in their female counterparts. This suggests that *Spiroplasma* is either able to maintain infection during metamorphosis, possibly due to extracellular proliferation (Goto et al., 2006), or that it can rapidly re-colonize upon reformation of the gut. *Spiroplasma* density was also significantly higher in the testes of teneral males than in the ovaries of teneral females. Localization to the testes suggests that *Spiroplasma* may be sexually transmitted from males to females, as has already been observed with *Sodalis* in tsetse flies. The above properties can be exploited in paratransgenic approaches in a similar way to those currently being explored for *Sodalis* (De Vooght et al., 2015).

In a collapsing colony of *G. f. fuscipes* flies, live females had a higher *Spiroplasma* density than prematurely dead females. This was true of both gravid and non-gravid females, and indicates that *Spiroplasma* may alter adult female fitness. No difference in *Spiroplasma* density was observed between live and dead males from the same colony. It is therefore possible that *Spiroplasma* could play a protective role, as has been observed in other facultative strains of *Spiroplasma* (Xie et al., 2010; Paredes et al., 2015).

5.5.4 Whole genome shotgun sequencing of novel taxa

Whole genome shotgun sequencing confirms the presence of several novel taxa at high relative abundance in multiple species of tsetse fly that were detected by 16S rRNA gene amplicon profiling. *Spiroplasma* and *Klebsiella*

have not previously been isolated from tsetse flies by culture dependent or independent methods. However, they are both commonly associated with arthropods, where they are capable of inducing a number of phenotypes from pathogenicity to mutualism (Bolanos et al., 2015; Dillon and Dillon, 2004). *Sodalis* prevalence is well documented in tsetse flies. Despite this, its role within the host is not well understood. In combination with existing genomic resources, the draft genomes presented here may help us to understand the increased prevalence and relative abundance of *Sodalis* in *pallidipes* and *fusca* group flies (Aksoy et al., 2014; Dennis et al., 2014), and perhaps their function within the host. The distribution and genomic characteristics of some novel taxa observed in this study suggest that they may have been associated with their host for a considerable amount of time. This should be taken in to account when planning future studies in to the prevalence and abundance of different members of the microbiome.

A *Rickettsia*-Like Organism (RLO) was one of the first bacteria, excluding *Wigglesworthia*, to be identified in association with the tsetse fly (Reinhardt et al., 1972) . RLOs were identified in multiple tsetse species (*G. brevipalpis*, *G. f. fuscipes*, G*. m. morsitans*, and *G. pallidipes*), and localized intracellularly to several locations within the host (midgut, bacteriome, oocytes, and fatbody) (Reinhardt et al., 1972). Maudlin and Ellis (Maudlin and Ellis, 1985) proposed a positive correlation between RLO infection and susceptibility to trypanosome infection. Welburn *et al.* suggested that the mechanism reinforcing the interaction was endochitinase activity by RLOs, which allowed trypanosomes to breach the peritrophic membrane of the gut (Welburn and Maudlin, 1991). *Sodalis* was subsequently isolated from *G. m. morsitans* (Dale and Maudlin, 1999), and was presumed to be the RLO (Welburn and Maudlin, 1999), though the reasoning behind this was not obvious, or evidenced, given the phylogenetic distance of the two organisms. In recent years Mediannikov *et al.* (Mediannikov et al., 2012) have rediscovered a member of the *Rickettsiaceae* in the midgut and hemocytes of *G. m. submorsitans* flies. This finding was compounded here in the same species of tsetse fly. The generation of genomic resources for both *Sodalis* and *Rickettsia* provides the opportunity to investigate the biology and phylogenetic

history of both organisms in future studies, and to solve the mystery of their shared identity.

Through targeting novel taxa with metagenomic sequencing we have obtained genomic resources for stably co-infecting bacteria such as *Wigglesworthia*, *Sodalis* and *Wolbachia*, along with large portions of the host genome. This lends huge potential to future comparative genomic studies, and enables a community-wide approach to studying microbial interactions within the tsetse fly. Areas of particular interest are: the dissemination of functional traits amongst the microbiota; how host ecology influences microbial infection frequency and functional repertoire; and the frequency of co-infection with, or mutual exclusion by, multiple potential reproductive manipulators (e.g. *Wolbachia*, *Spiroplasma*, and *Rickettsia*). Deep sequencing has exposed the hidden diversity of bacterial communities associated with tsetse flies, and with it the potential for new control strategies that harness novel tsetse biology.

Chapter 6 : General Discussion

6.1 Summary of the field

Symbiotic associations between bacteria and insects are widespread and diverse in their ecology and evolutionary dynamics (Buchner, 1965; Douglas, 2010). They elicit a broad range of phenotypes from their hosts and can achieve transmission through horizontal, vertical, intracellular and extracellular transmission mechanisms (Engel and Moran, 2013; Salem et al., 2015a). In addition to their impact on host fitness, bacterial partners can influence the biology of other members of the communities that live within organisms (Hussa and Goodrich-Blair, 2013). Due to the multipartite, intricate and often obligate nature of these associations, they can be difficult to dissect experimentally (Garcia and Gerardo, 2014; Mushegian and Ebert, 2016). However, bacterial mutualists offer novel approaches to the control of important insect agricultural pests and human and animal pathogen vectors (Beard et al., 2002). The advent of next generation sequencing has been of huge advantage to the field of symbiosis research, and in combination with field and laboratory-based techniques, has revolutionised the way in which we are able to analyse complex symbiotic interactions (McCutcheon and Moran, 2012). Our increasing knowledge of how symbionts interact with their hosts has already generated several novel control methods, and shows great promise for more effective and environmentally friendly insect control strategies in the future (Douglas, 2007; Berasategui et al., 2016).

6.2 Bacteria associated with the olive fruit fly

6.2.1 Ecological conditions alter community composition

Previous studies of the *B. oleae* microbiota have focused upon *Ca.* E. dacicola due to its high abundance and infection frequency in natural populations (Capuzzo et al., 2005; Sacchetti et al., 2008; Ben-Yosef et al., 2015). However, *Ca.* E. dacicola is not ubiquitous in any population studied to date, and all laboratory colonies of *B. oleae* are able to survive in its absence, indicating that the mutualism is condition-dependent and has not yet reached fixation through mutual obligacy (Estes et al., 2012a). I used high-throughput 16S rRNA gene amplicon sequencing and qPCR to show that the structure and composition of bacterial communities associated with adult *B. oleae* vary when host ecology changes either through seasonal fluctuations, or simply by changing laboratory environments (Chapter 2). Instability in the composition of the microbiota is reminiscent of the high turnover of gut microbes in *D. melanogaster*, which display high variation in community composition between natural and laboratory populations (Wong et al., 2013). Similarly, diet shapes the composition of the microbiota in this system, as it is presumed to in *B. oleae*, but is not the sole determinant of gut colonisation (Chandler et al., 2011).

As well as ensuring the inheritance of beneficial metabolic traits provided by microbes, vertical transmission can serve to protect vulnerable, nutrient-rich niches such as the gut lumen from opportunistic pathogens by ensuring their occupation with mutualists (Johnston and Rolff, 2015; Salem et al., 2015b). In *B. oleae*, perturbing the microbiota results in health deficits in adults (Chapter 4). This is true even of laboratory-reared flies, which do not benefit from an evolutionarily adapted gut microbiota, indicating that collateral benefits of the native microbiota should be explored when trying to improve mass-rearing strategies.

6.2.2 *Ca.* E. dacicola as a specialised gut symbiont

There are only a few examples of highly specialised, co-evolving, extracellular gut symbionts of insects (Salem et al., 2015a). Specialisation here is classified as the co-adaptation of host and symbiont towards maintenance of the partnership (Kwong and Moran, 2015). The evolution of specialisation in gut bacteria may be prevented by the increased heterogeneity of the gut environment, more frequent opportunities for genetic exchange with other organisms, and less stringent host regulation of physiological and metabolic conditions (Moran, 2002). In addition, gut-associated bacteria can be both vertically and horizontally transmitted, increasing the potential number of hosts, and reducing the evolutionary requirement for specialisation (Kwong and Moran, 2015). Despite these barriers, specialised gut symbioses have evolved in some insects, and they often exhibit similar genomic traits to those seen in intracellular endosymbionts. This is due to the similar population bottlenecks experienced by both, as gut symbionts also adopt specialised mechanisms for vertical transmission between host generations (Hosokawa et al., 2006; Kikuchi et al., 2009; Kaiwa et al., 2014).

Three traits are thought to determine the propensity of gut microbes to become specialised: mode of transmission, net fitness benefit to the host, and net fitness benefit to the microbe (Kwong and Moran, 2015). *Ca.* E. dacicola exhibits the traits of a specialised extracellular gut bacterium, as it is vertically transmitted from mother to offspring by egg-smearing (Estes et al., 2009), has a reduced genome (Chapter 3), and is thought to contribute to host fitness (Chapter 3) (Ben-Yosef et al., 2010; 2014; 2015). *Ca.* E. dacicola is non-culture-viable, and is only found in association with its *B. oleae* host, suggesting that it obtains some fitness benefit from the association (Chapter 3) (Estes, 2009). This has never been tested empirically, and is a fairly novel area of research in the symbiosis field due to the logistical difficulties of studying obligate microbes (Garcia and Gerardo, 2014; Mushegian and Ebert, 2016). While the *Ca.* E. dacicola draft genome is reduced in size, it is not reduced in GC content in comparison to free-living bacteria (Chapter 3), as was predicted by Estes *et al.* from several protein-coding genes (Estes, 2009). AT-bias in endosymbiont genomes results from the random deletion of

DNA repair genes and subsequent fixation of mutations that cannot be purged by purifying selection due to small effective population size (Parker and Marinus, 1992; Moran, 2002; Moran et al., 2009). In free-living bacteria codon usage is optimised for translational efficiency by the availability of tRNA genes and genome size (Reis, 2004). The fact that GC content is maintained in *Ca.* E. dacicola despite genome erosion indicates that this organism is still under selective pressure to maintain replication efficiency. This may be due to competition from other microbes within the gut, or from the host itself.

6.2.3 Facultative symbionts

Observational and genomic data suggest that a novel bacterium, *Tatumella* TA1, is stably associated with natural populations of *B. oleae* in both Israel and Crete, and that it may be able to contribute to *B. oleae* fitness through nutritional interactions (Chapter 2, Chapter 3). Estes *et al.* identified another potential facultative symbiont in wild populations of *B. oleae* collected in the USA (Estes et al., 2009). An *Enterobacter* sp. was detected by PCR and culturing techniques, and similarly to *Tatumella* TA1 was isolated from all life stages of *B. oleae*, maintaining infection throughout metamorphosis. Genomic data is not yet available for this *Enterobacter* sp., but it was not detected by 16S rRNA gene studies of any of the wild populations in this thesis (Chapter 2). Variation in the geographic distribution of multiple secondary symbionts has been identified previously in pea aphids, and is determined by environmental conditions due to the contribution of facultative symbionts to host fitness in a condition-dependent manner (Tsuchida et al., 2002; Russell and Moran, 2006). In order to better understand how the microbiota contributes to *B. oleae* fitness in response to ecological conditions, interactions between bacterial partners must first be investigated. It is not yet known whether *Tatumella* TA1 or *Enterobacter* sp. interact with *B. oleae, Ca.* E. dacicola or both, or whether these interactions are mutualistic or antagonistic.

6.2.4 Conflict and cooperation in gut microbes

Due to their extracellular nature, the evolution of gut symbionts is shaped not just by interactions with the host, but also with other bacteria (Mushegian and Ebert, 2016). In humans, syntrophy promotes cooperation between some members of the gut microbiota: *Bacteroides ovatus* acts as a keystone member of the community by providing secreted enzymes for the digestion of polysaccharides that are not required for its own growth, but which provide other members of the gut microbiota with essential nutrients (Rakoff-Nahoum et al., 2016). These species reciprocally provide nutritional resources, which lead to a net fitness benefit for *B. ovatus*, and promote the maintenance of an ecologically diverse gut microbiota. Public goods have also been identified as a source of conflict, rather than cooperation, amongst microbes (Xavier, 2011). Whatever their net effect, both strategies require some form of recognition mechanism to enable exchange between partners. Adhesion has been hypothesised as a key host determinant of microbial community structure in gut environments, through host recognition and bacterial-bacterial contact (O'Toole et al., 2000; McLoughlin et al., 2016). The discovery of large secreted adhesion protein genes in *Tatumella* TA1 suggests that either it plays a role in host cell adhesion, or is able to exploit host recognition of *Ca.* E. dacicola. Functional genomic analyses indicate that *Tatumella* TA1 and *Ca*. E. dacicola show similar adaptations to the insect gut environment, but it is not yet known whether they occupy the same ecological niche within *B. oleae* (Chapter 3). Assessing how microbes interact within the *B. oleae* gut, including those that are stably and transiently associated is a key step in deciphering how they contribute to host fitness, and how they interact with each other. This is of critical importance to the *B. oleae* microbiota due to its susceptibility to perturbation (Chapter 4).

6.2.5 The role of symbiotic bacteria in nutrition

There is a large body of evidence that the *B. oleae* gut is colonised by other microbes alongside, and sometimes in the place of, *Ca.* E. dacicola (Kounatidis et al., 2009; Estes et al., 2012b; Chapter 2,; Chapter 3). However, all documented phenotypic effects of the microbiota in wild *B. oleae* have

been attributed to *Ca.* E. dacicola, even though the resulting phenotypes are induced by manipulation of the entire community. This is of critical importance in this system in particular, as unlike bacteriocyte-associated symbionts, *Ca.* E. dacicola ecology and evolution is shaped not just by interactions with the host, but also with other bacteria (Mushegian and Ebert, 2016). I have shown that both *Ca.* E. dacicola and *Tatumella* TA1 encode complete pathways for the synthesis of the essential amino acid lysine, which is lacking in the larval diet (Ben-Yosef et al., 2015; Chapter 3). In addition, all steps for the assimilation of ammonia from uric acid are encoded in a mosaic fashion between both organisms, excluding the enzyme allantoicase (Chapter 3). Allantoicase was not detected in the host or any member of the microbiota, but is a key step in the assimilation of ammonia from uric acid (Chapter 3). Fungal members of the gut microbiota are potential candidates for the origin of this enzyme, due to their role in nitrogen recycling in other insects (Sasaki et al., 1996), that were not studied in this thesis.

6.3 Future work

The data presented in this thesis suggest that the *B. oleae* microbiota is susceptible to perturbation under fluctuating environmental conditions, and that alterations in community composition could have ramifications for host metabolism and health. However, many questions remain regarding the underlying biology of the symbiosis, and how knowledge of these characteristics may be applied to control strategies.

6.3.1 Transmission Mechanisms

Some members of the bacterial community are present at all life stages in both natural and laboratory populations, implying vertical transmission (Chapter 2, Chapter 4). It is not yet clear whether adapted (natural) and non-adapted (laboratory) microbes employ the same transmission mechanisms, whether these are vertical, horizontal, or mixed-mode, or whether the host facilitates transmission in any way. This is an important factor in determining the role of opportunistic pathogens in host biology, both inside and outside of the laboratory environment. Determining transmission mechanisms will help us to better understand the within-community dynamics of the microbiota, and the ancillary effects of microbes on host fitness such as competitive exclusion.

6.3.2 Specificity of Interactions

The interaction between *B. oleae* and *Ca*. E. dacicola is well established, and genomic analyses indicate that *Ca.* E. dacicola is host-adapted and dependent upon vertical transmission (Chapter 3). However, the non-ubiquitous and condition-dependent nature of the symbiosis is more reminiscent of a facultative interaction that is evolving towards obligacy than a fixed mutualism. This represents a fairly risky evolutionary strategy for *Ca.* E. dacicola, as there is no evidence that it can replicate *ex vivo*, indicating that it has reached an evolutionary dead-end and is entirely dependent upon host fitness for survival. In addition, despite its obligate nature, it is not currently clear how *Ca.* E. dacicola benefits from the interaction with *B. oleae*. The story is made more complex still by the evidence that *Tatumella* TA1 may

participate in some of the metabolic functions that are essential to host development that are currently attributed to *Ca.* E. dacicola (Chapter 3). Further research is required to determine the inter-related dependencies and conflicts between partners, and the evolutionary stability of these interactions. The condition-dependent nature of the *B. oleae* microbiota is a powerful tool in this endeavour. Transcriptomic and metabolomic analyses of host-microbiota metabolic exchanges under varying ecological conditions may reveal the selective pressures that reinforce the presence or absence of certain members, and highlight fitness conflicts and niche partitioning between partners.

6.3.3 Seasonal community dynamics

Both insect and bacterial growth is temperature-dependent, and is expected to vary across seasons. In addition, both the juvenile and adult *B. oleae* diet varies seasonally in composition and nutritional quality. Despite seeing annual variation in community composition (Chapter 2), we do not currently know how variation in specific environmental conditions such as temperature and food availability affects the community dynamics of bacteria associated with *B. oleae*. This could be addressed by using 16S rRNA gene amplicon profiling or qPCR to track community composition over multiple time points in the same population, to determine how physiological conditions and host mechanisms interact to control microbiota community dynamics, and what impact this has upon the fitness of the symbiosis.

6.3.4 Manipulating the microbiota for mass rearing

Tatumella TA1 has been identified as a potential probiotic candidate for *B. oleae* mass rearing. It has a putative role in nitrogen provisioning in natural populations, and could be used to improve diet composition and utilisation in laboratory-reared populations. In addition, its apparent adaptation to the physiological conditions in the gut and ability to be vertically transmitted may enable it to adopt a defensive role against opportunistic pathogens in the laboratory environment through niche occupation. Work in this thesis has highlighted the susceptibility of the *B. oleae* gut environment to perturbation,

and direct or indirect interactions of adapted microbes with opportunistic pathogens may serve to significantly increase the survival of laboratory insects.

6.4 Bacteria associated with the tsetse fly

6.4.1 Communities are not restricted to *Wigglesworthia*, *Sodalis* and *Wolbachia*

16S rRNA gene profiling experiments revealed much higher bacterial community diversity in all tsetse species than in previous culture-independent studies, and the presence of multiple taxa that have been characterised as endosymbionts and reproductive manipulators in other insects (Chapter 5). The diversity discovered was surprising, as the exposure of tsetse to environmental microbes is thought to be limited due to their intrauterine development, reliance upon sterile blood as a nutrient source, and the vertical transmission of endosymbionts that occupy ecological niches within the fly from an early age.

6.4.2 Sodalis glossinidius

Sodalis infections were more frequent and at higher abundance in *G. medicorum* flies than in any other population studied. This reflects a general increase in *Sodalis* infection rate in *fusca* group flies (Dennis et al., 2014). The genomic data generated for two *G. medicorum* individuals alongside their *Sodalis* and *Wigglesworthia* isolates provides an opportunity to scrutinize this multipartite interaction for genotype-genotype factors that may promote colonization and maintenance. These interactions are features that are of importance to paratransgenic approaches to trypanosome control as they are integral to *Sodalis* infectivity (De Vooght et al., 2014).

6.4.3 Reproductive manipulators

Rickettsia and *Spiroplasma* have both been characterised as reproductive manipulators in several insect species (Duron et al., 2008). Conspecific reproductive manipulators display competitive exclusion in some insects where they share tissue tropism and overlaps in ecological niches (Hughes et al., 2014). The discovery of two novel potential reproductive manipulators alongside previously characterised *Wolbachia* infections suggests that bacteria-bacteria interactions may play an important role in determining

endosymbiont infection rates in tsetse flies, which in turn may cause significant variation in vector biology. *Rickettsia*, *Sodalis* and *Spiroplasma* infections were all restricted to or elevated in one tsetse fly subgroup. This has previously been observed in *Wolbachia* infections, which are restricted to, but not ubiquitous amongst, the *fusca* and *morsitans* subgroups (Doudoumis et al., 2012). Localisation of *Spiroplasma* to the adult reproductive tissues suggests that it may share an ecological niche with *Wolbachia*, and that the evolutionary age of the relationship may explain their mutual exclusion in tsetse subgroups. The presence of *Spiroplasma* symbionts may impede the success of *Wolbachia*-based control strategies. Therefore, their ecological overlap warrants further investigation.

6.5 Future work

The data presented in this thesis suggest that there are several reproductive manipulators and bacterial taxa of ecological significance circulating in tsetse fly populations, and that their presence and phenotype may be determined by host biology and other bacteria. Further characterisation of these additional taxa has the potential to provide a great deal of information on the role of multipartite interactions on vector biology.

Spiroplasma

Genotypic variation in *Spiroplasma* strains isolated from *G. tachinoides* and *G. f. fuscipes* individuals suggests that either there are different strains of *Spiroplasma* circulating locally in distinct populations, or that *Spiroplasma* is co-evolving with the host due to fidelitous vertical transmission. Further MLST analysis of *Spiroplasma* strains originating from geographically and genotypically distinct populations will help to determine *Spiroplasma* population dynamics in tsetse flies. In addition, studies of the ecology of *Spiroplasma* strains through genotypic and transcriptomic analyses will determine the potential cross-over in ecological niches between *Spiroplasma* and *Wolbachia*, and the role that *Spiroplasma* might play in host biology.

Sodalis

Sodalis titre is proposed to be regulated in part by the availability of nutrients and crossfeeding with the primary endosymbiont *Wigglesworthia* (Snyder et al., 2010). Increased *Sodalis* prevalence and relative abundance in flies of the *fusca* subgroup suggest that these mechanisms are differentially regulated or evaded by *Sodalis* in this host genotype background. Due to the co-speciation of *Wigglesworthia* with the host, and differences in evolutionary trajectories between tsetse species, it is possible that adaptations in the host, *Wigglesworthia*, *Sodalis*, or a combination thereof could be responsible for the difference in *Sodalis* titre between tsetse species. Other members of the interaction that interfere with nutrition, such as trypanosomes, may also play a role (Telleria et al., 2014). Comparative genomic analysis of the metagenomic

data generated for two *G. medicorum* flies and their associated *Wigglesworthia* and *Sodalis* strains with individuals from species that were not infected with *Sodalis* may shed some light on the role of host-microbiota genotypes in increased *Sodalis* prevalence.

6.6 General Conclusions

Despite the presence of obligate symbionts in both systems studied in this thesis, the importance of studying associated microbes as consortia and incorporating even minor members of the community in to experimental frameworks is evident. Profiling approaches revealed surprising plasticity in community composition, emphasising the importance of ecological and evolutionary context in studying host-microbe interactions. Most strikingly, functional analysis reinforces the importance of considering not just host-microbe interactions, but also microbe-microbe interactions. Both systems lend themselves to future studies of conflict and cooperation between microbes, and how these interactions affect host fitness and the evolutionary trajectories of insect-microbe symbioses.

Publications

Blow F, Gioti A, Starns D, Ben-Yosef M, Pasternak Z, Jurkevitch E, Vontas J, Darby AC. (2016). Draft genome sequence of the *Bactrocera oleae* symbiont "*Candidatus* Erwinia dacicola". Genome Announc 4: e00896-16.

Blow F, Vontas J, Darby AC. (2016). Draft genome sequence of *Stenotrophomonas maltophilia* SBo1 isolated from *Bactrocera oleae*. Genome Announc. 4: e00905-16.

Blow F, Vontas J, Darby AC. (Submitted). Draft genome sequence of *Chryseobacterium* CBo1 isolated from *Bactrocera oleae*. Genome Announc.

Blow F, Vontas J, Darby AC. (In Preparation). Draft genome sequence of *Tatumella* TA1 isolated from *Bactrocera oleae*. Genome Announc.

Doudoumis V*, **Blow F***, Saridaki A, Augustinos A, Dyer N, Goodhead I, Solano P, Rayaisse JB, Takac P, Mekonnen S, Parker A, Darby A, Bourtzis K, Tsiamis G. (Submitted). Challenging the Wigglesworthia, Sodalis, Wolbachia symbiosis dogma in tsetse flies: Spiroplasma is present in both laboratory and natural populations.
*These authors contributed equally.

Chapter 7 : Appendices

7.1 Appendix 1

7.1.1 *B. oleae* artificial rearing practices

Laboratory colonies of *B. oleae* were reared following standard small-scale rearing practices (Genc and Nation, 2008). All rearing occurred in a sterile laboratory maintained at 25 ^{o}C and 60 % relative humidity. Adults were housed in 30 cm^{3} bug dorm cages and supplied with ceresin wax cones for laying. 100 g adult artificial diet consisted of 10 g hydrolysed yeast, 75 g icing sugar, and 3 g egg yolk, and was supplied in a clean petri dish lid on a daily basis. Sterile Milli-Q water was provided in a clean plastic pot with absorbent paper towels. Eggs were collected from wax cones by washing the inside of the cone with sterile Milli-Q water on a daily basis. Eggs were washed on to moistened filter paper and incubated for 48 h in a plastic container lined with moistened paper towels. One litre of larval diet consisted of 250 g cellulose powder, 30 g soy hydrolysate, 20 g sucrose, 75 g Brewer's yeast, 2 g nipagen, 4.5 ml 10M HCl, 20 ml olive oil, 7.5 ml tween 80, and 0.5 g potassium sorbate. Approximately 200 first instar larvae were transferred to each petri dish filled with larval diet. Larval petri dishes were then sealed and incubated for 10 days in clear plastic containers lined with sterilised sand and holes in the lids for oxygenation. On the eighth day of development the petri dish lids were removed to allow wandering larvae to exit the diet and crawl in to the sand to pupate. On the eighth day of pupation larval food dishes were removed from the boxes and pupae were collected by sieving the sand. Pupae were placed in to clean petri dishes and placed in adult Perspex cages ready for emergence. Following this system, each generation was reared over a period of approximately four weeks.

7.2 Appendix 2

7.2.1 Chapter 3 supplementary information

Table 7.1 | *Ca. E. dacicola* assembly statistics. Statistics are presented for individual single cell and boil preparation libraries, combined assemblies, and hybrid assemblies with metagenomic shotgun data for *Ca. E. dacicola*. SC (Single Cell); BP (Boil Preparation); GSG (Green Shotgun from larvae); MP (Mate Pair from larvae).

Assembly Name	Number of Contigs	Total Length	Largest Contig	N50	GC (%)	% Complete Free Living	% Complete Endosymbiont	Library Type
FP	1447	1,729,198	29,211	2256	52.08	65.2	70.9	SC
C6	39	63,797	12,885	8994	50.66	0.0	0.0	SC
BP	19	14,221	5216	5216	45.35	0.0	0.0	BP
AH1	8	8546	5513	5513	44.87	0.0	0.0	BP
C3	215	189,261	33,113	1685	48.66	2.9	3.1	SC
C8	785	785,277	29,632	2355	50.59	27.5	29.9	SC
AG1	659	531,407	20,987	1895	49.3	8.0	8.7	BP
AG2	868	676,004	32,183	2169	49.73	10.9	11.8	BP
SC BP	2381	2,604,917	29,632	3061	51.87	81.9	89.0	SC BP
Hybrid	1239	3,388,615	181,380	8457	52.58	97.1	105.5	SC BP GSG MP
ErwSC	333	2,132,378	37,608	9998	53.52	92.0	100.0	SC BP GSG MP

7.3 Appendix 3

7.3.1 Chapter 5 supplementary information

Table 7.2 | Primers used for PCR, qPCR and sequencing reactions (SR).

Name	Gene region	Annealing temperature / extension time	Fragment size (bp)	Primer type	References
12SCFR 12SCRR	*mt 12S rRNA host*	54 °C / 1min	377	PCR	Hanner & Fugate 1997 Hanner & Fugate 1997
63F TKSSsp	*16S rRNA Spiroplasma*	59 °C / 1min	455	PCR, SR	Mateos et al, 2006 Fukatsu and Nikoh 2000
16SA1 Rick 16SR	*16S rRNA Rickettsia*	55 ºC / 1 min	200	PCR	Fukatsu and Nikoh 1998 Fukatsu et al. 2001
CLO f1 CLO r1	*16S rRNA Cardinium*	56 ºC / 1:30min	466	PCR	Gotoh et al. 2007
CLOF CLOR	*16S rRNA Cardinium*	54 ºC / 1:30min	450	PCR	Weeks et al. 2003
ArsF ArsR2	*16S rRNA Arsenophonus*	60 ºC / 1:30min	800	PCR	Duron et al. 2008
16STF1 TKSSsp	*16S rRNA Spiroplasma*	56 °C / 30sec	94	qPCR	Haselkorn et al. 2009 Fukatsu and Nikoh 2000
GmmtubqF GmmtubqR	host *β-tubulin*	56 °C / 30sec	151	qPCR	Guz et al. 2007 Guz et al. 2007
FqdnaA RqdnaADoud	dnaA	56 oC / 30sec	138	qPCR	Harumoto et al. 2014 This study
63F 16STR1_Hasel	*16S rRNA Spiroplasma*	67 ºC/ 1:30min	1334	PCR, SR	Fukatsu & Nikoh 2000 Haselkorn et al. 2009
16F2_Bi 23R1_Bi	region 16S-23S *Spiroplasma*	67 ºC / 1min	1024	PCR, SR	Bi et al. 2008 Bi et al. 2008
23F1_Bi 5R_Bi	region 23S-5S *Spiroplasma*	67 ºC / 3min	2914	PCR, SR	Bi et al. 2008 Bi et al. 2008
23F2_BiDoud	region 23S-5S *Spiroplasma*	-	-	SR	This study
23F3_Bi	region 23S-5S *Spiroplasma*	-	-	SR	Bi et al. 2008
23F3_Heres	region 23S-5S *Spiroplasma*	-	-	SR	Heres et al. 2010
23R2_Bi	region 23S-5S *Spiroplasma*	-	-	SR	Bi et al. 2008
fru-f fru-r	*fruR Spiroplasma*	56 ºC / 30sec	398	PCR, SR	Montenegro et al. 2005 Montenegro et al. 2005
SRdnaAF1 SRdnaAR1	*dnaA Spiroplasma*	50 ºC / 30sec	515	PCR, SR	Anbutsu & Fukatsu 2003 Anbutsu & Fukatsu 2003
FtsZF2 FtsZR3	*ftsZ Spiroplasma*	57 ºC / 1min	774	PCR, SR	Haselkorn et al. 2009 Haselkorn et al. 2009
ParEF2 ParER2	*pare Spiroplasma*	57 ºC / 1min	1126	PCR, SR	Haselkorn et al. 2009 Haselkorn et al. 2009
RpoBF1 RpoBR2	*rpoB Spiroplasma*	60 ºC / 1:30min	1703	PCR, SR	Haselkorn et al. 2009 Haselkorn et al. 2009
RpoBR4	*rpoB Spiroplasma*	-	-	SR	Haselkorn et al. 2009

Table 7.3 | Richness and diversity estimates of bacterial communities from laboratory populations of tsetse flies profiled by 16S rRNA gene amplicon sequencing. *Gmm*: *Glossina morsitans morsitans; Gff*: *Glossina fuscipes fuscipes; Gpal*: *Glossina pallidipes.*

		Species richness indices		Species diversity indices	
Samples	**Number of OTUs**	**Chao1**	**ACE**	**Shannon**	**Simpson**
Gmm larvae	139.0±36.0	84.4±30.7	90.9±31.5	1.73±0.23	0.64±0.08
Gmm gut male 1d	74.0±10.8	55.6±9.5	64.8±13.3	1.38±0.28	0.55±0.13
Gmm testes 1d	89.7±12.3	77.5±14.1	82.0±11.5	1.45±0.30	0.54±0.12
Gmm gut female 1d	53.3±6.9	32.8±5.6	35.5±6.4	1.14±0.15	0.44±0.07
Gmm ovaries 1d	121.0±16.3	86.8±15.0	92.3±6.3	1.57±0.38	0.53±0.11
Gmm gut male 15d	57.7±6.7	41.1±9.4	44.4±10.1	1.46±0.18	0.59±0.09
Gmm testes 15d	132.3±14.9	91.1±18.6	99.4±19.5	1.05±0.10	0.37±0.03
Gmm gut female	55.3±1.5	48.3±13.2	40.5±5.0	1.32±0.12	0.53±0.05
Gmm ovaries 15d	70.3±18.6	56.1±9.0	59.6±11.7	1.05±0.07	0.39±0.01
Gff larvae	63.7±1.8	64.9±8.2	62.1±12.5	1.45±0.26	0.59±0.11
Gff gut male 1d	39.3±1.7	33.6±2.7	33.3±1.0	0.97±0.06	0.37±0.03
Gff testes 1d	152.7±47.9	86.2±20.0	88.3±20.5	1.55±0.24	0.58±0.11
Gff gut female 1d	40.7±2.6	27.8±0.3	29.1±0.3	1.18±0.14	0.47±0.07
Gff ovaries 1d	139±7.2	148.0±5.9	150.4±6.1	2.2±0.013	0.8±0.15
Gff gut male 16d	45.0±4.7	26.2±1.2	26.1±0.9	1.48±0.01	0.61±0.00
Gff testes 16d	75.0±19.0	82.6±22.3	73.6±17.1	1.16±0.13	0.43±0.07
Gff gut female 16d	41.7±3.5	42.5±8.3	34.2±3.3	1.39±0.08	0.58±0.04
Gff ovaries 16d	103.3±12.2	95.3±15.3	92.1±10.9	0.95±0.07	0.35±0.02
Gpal larvae	92.7±20.7	63.7±22.1	59.1±12.5	0.98±0.06	0.37±0.02
Gpal gut male 1d	46.7±6.2	18.4±0.5	19.5±0.4	0.75±0.03	0.34±0.01
Gpal testes 1d	120.5±13.5	104.3±2.7	98.7±0.4	2.00±0.53	0.64±0.17
Gpal gut female 1d	32.5±2.9	21.6±4.0	21.7±3.3	0.69±0.00	0.35±0.01
Gpal ovaries 1d	309.7±21.2	180.4±16.9	192.9±11.5	2.62±0.08	0.82±0.01
Gpal gut male 16d	42.3±14.5	28.2±2.9	29.8±1.7	1.36±0.15	0.54±0.06
Gpal testes 16d	155.7±12.3	102.8±14.6	116.9±22.5	1.61±0.10	0.55±0.04
Gpal gut female 16d	43.7±6.7	29.0±4.1	29.1±3.0	1.24±0.08	0.52±0.02
Gpal ovaries 16d	131.7±6.5	93.7±11.2	100.5±10.5	1.59±0.09	0.57±0.04

Table 7.4 | Richness and diversity estimation of the 16S rRNA libraries from the amplicon sequence analysis of the natural populations. *Gmm*: *Glossina morsitans morsitans; Gff*: *Glossina fuscipes fuscipes; Gpal*: *Glossina pallidipes*; *Gmed: Glossina medicorum; Gmms: Glossina morsitants. Submorsitans*; *Gtach: Glossina tachinoides.* WI: whole insect

		Species richness indices		Species diversity indices	
Samples	**Number of OTUs**	**Chao1**	**ACE**	**Shannon**	**Simpson**
Gff Busime Female Gut	106.66±20.70	158.07±35.08	159.47± 30.17	0.60±0.07	0.88±0.02
Gff Busime Male Gut	130.25±16.68	182.59±27.12	187.56± 33.36	0.67±0.08	0.86±0.02
Gff Dokolo Female Gut	146.57±33.61	202.42±47.58	200.96±41.47	0.61±0.11	0.87±0.03
Gff Dokolo Male Gut	180.83±28.05	258.40±46.40	249.66±50.22	0.68±0.08	0.86±0.02
Gff Kaberamaido Female Gut	143.27±46.44	199.10±72.14	204.39±64.26	0.66±0.22	0.87±0.04
Gff Kaberamaido Male Gut	160.79±50.70	233.86±68.19	225.75±56.82	0.64±0.10	0.87±0.03
Gff Murchison Falls Female Gut	132±16.97	183.79±10.40	181.69±12.12	0.53±0.09	0.89±0.02
Gff Murchison Falls Male Gut	142.55±30.25	207.82±33.19	197.31±28.34	0.56±0.10	0.88±0.02
Gff Otuboi Female Gut	137.25±31.76	202.96±39.03	201.71±34.21	0.54±0.08	0.89±0.02
Gff Otuboi Male Gut	146±34.11	219.55±58.64	214.50±46.68	0.61±0.10	0.87±0.02
Gmed Burkina Faso Female [WI]	58.43±70.47	94.41±109.87	100.84±108.9	1.27±0.37	0.68±0.11
Gmed Burkina Faso Male [WI]	41±15.54	55.66±18.36	64.87±25.47	1.23±0.16	0.72±0.05
Gmm Murchison Falls Female Gut	78.5±7.94	115.48± 28.47	124.98±29.09	0.21±0.04	0.97±0.008
Gmm Murchison Falls Male Gut	97±28.28	128.53±44.58	136.91±48.21	0.24±0.04	0.96±0.006
Gmms Murchison Whole Insects	99.63±152.44	139.48±172.48	143.16±165.8	1.13±0.25	0.71±0.10
Gpal Murchison Falls Female Gut	92.5±17.72	123.1±21.71	123.43±18.29	0.68±0.54	0.83±0.14
Gpal Murchison Falls Male Gut	101.4±19.73	124.75±27.48	126.74±23.13	0.91±0.94	0.78±0.23
Gpal Burkina Faso Whole Insects	393.88±41.79	494.73±40.72	491.31±41.21	0.83±0.27	0.79±0.11
Gtach Burkina Faso Female [WI]	393.4±65.3	480.7±70.4	472.52±59.85	0.59±0.26	0.88±0.09

Table 7.5 | *Spiroplasma* MLST genotyping in *Glossina* species. Numbers in brackets indicate the number of polymorphisms observed.

	16S rRNA	*ITS*	*rpoB*	*dnaA*	*parE*	*fruR*	*ftsZ*
G. f. fuscipes Bratislava lab-colony ♀	Allele 1	Allele2	Allele3	Allele4	Allele5	Allele6	Allele7
G. f. fuscipes Bratislava lab-colony ♂	Allele 1	Allele2	Allele3	Allele4	Allele5	Allele6	Allele7
G. f. fuscipes Uganda ♀	Allele 1	Allele2	Allele3	Allele4	Allele5	Allele6	Allele7
G. f. fuscipes Seibersdorf lab-colony ♀	Allele 1	Allele2	Allele3	Allele4	Allele5	Allele6	Allele7
G. f. fuscipes Seibersdorf lab-colony ♂	Allele 1	Allele2	Allele3	Allele4	Allele5	Allele6	Allele7
G. tachinoides Burkina Faso	Allele 8 (4)	Allele9 (7)	Allelel0 (2)	Allele11 (1)	Allele12 (8)	Allele13 (1)	Allele14 (1)

Maximum Likelihood (ML) phylogenies were also constructed for Figure 7.1- Figure 7.6, and in all cases they displayed a similar topology to Bayesian trees. Therefore, only Bayesian phylogenies are presented here. Bayesian posterior probabilities and ML bootstrap values based on 1000 replicates are given at each node (only values >50 % are indicated). Asterisks indicate support values lower than 50 %. The *Spiroplasma* strains present in *G. f. fuscipes* and *G. tachinoides* are indicated in bold letters.

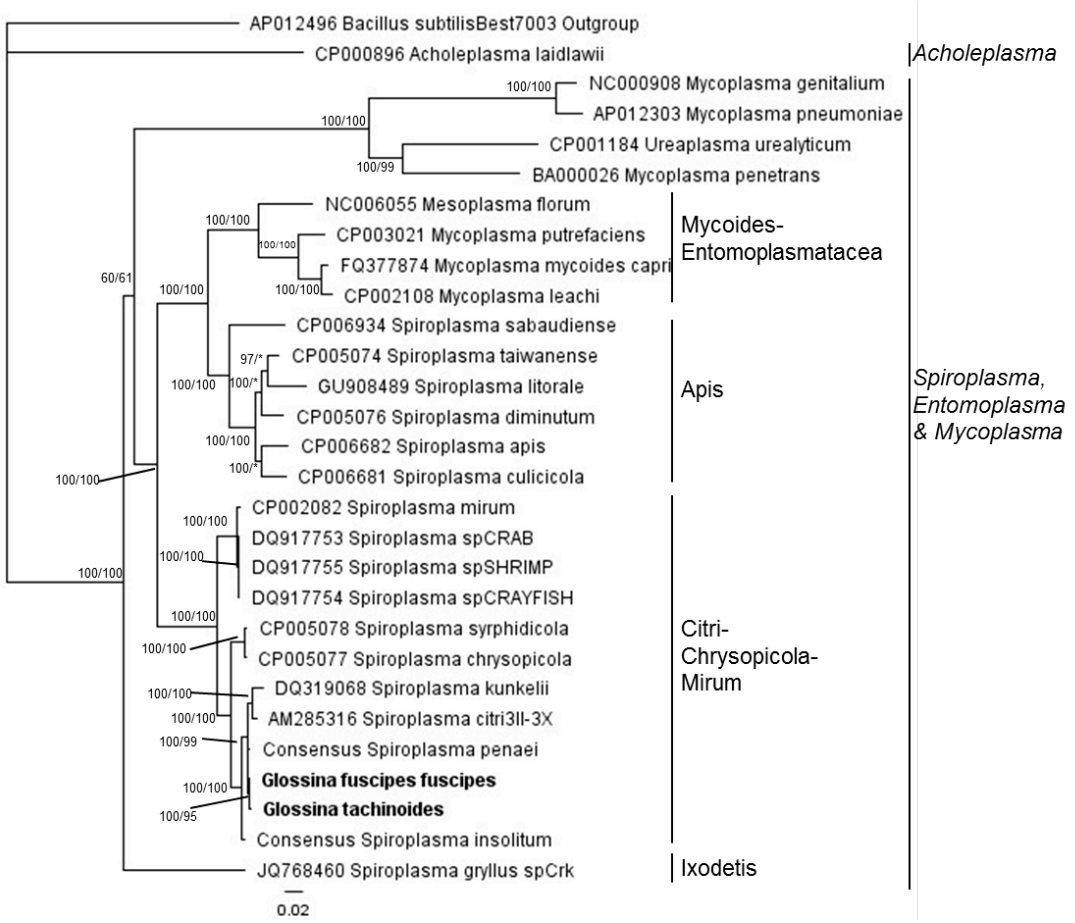


Figure 7.1 | Bayesian inference phylogeny of the 16S-23S-5S rRNA (ITS) region.

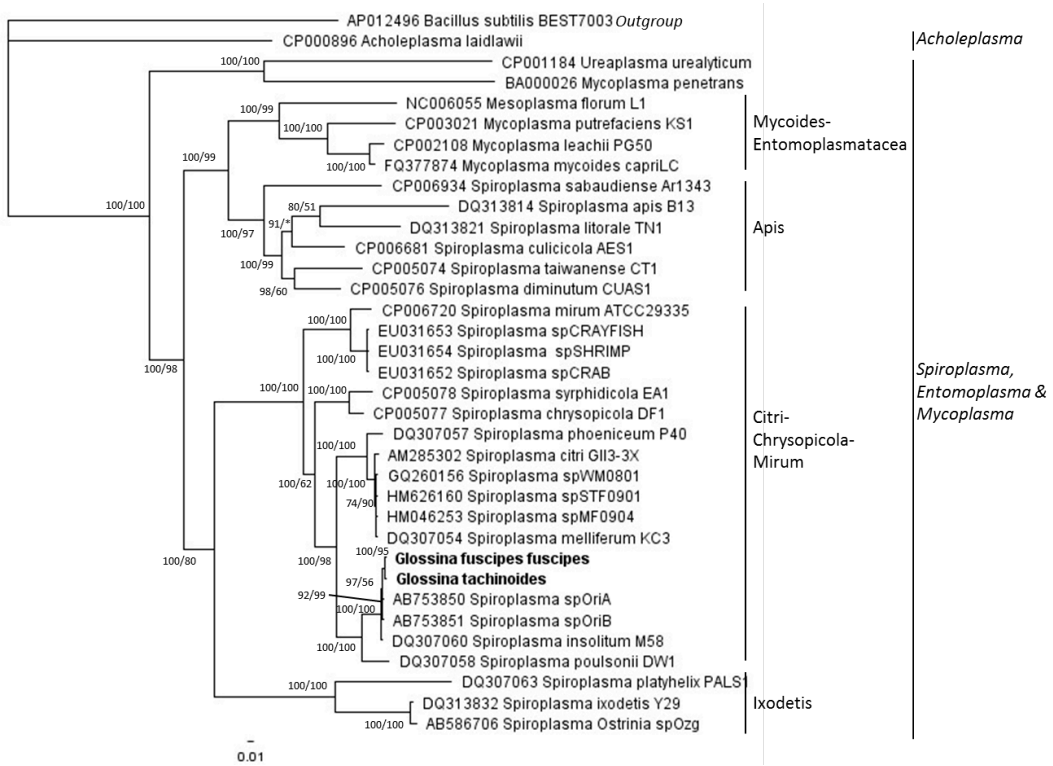


Figure 7.2 | Bayesian inference phylogeny of the *rpoB* gene.

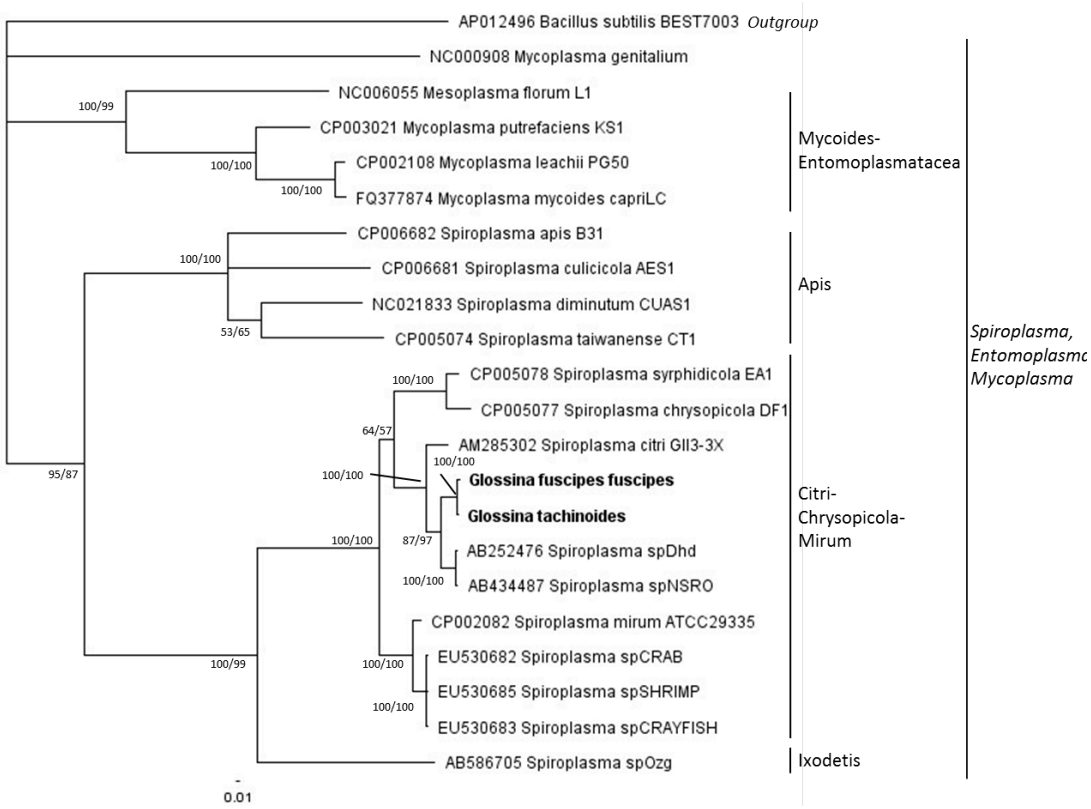


Figure 7.3 | Bayesian inference phylogeny of the *dnaA* gene.

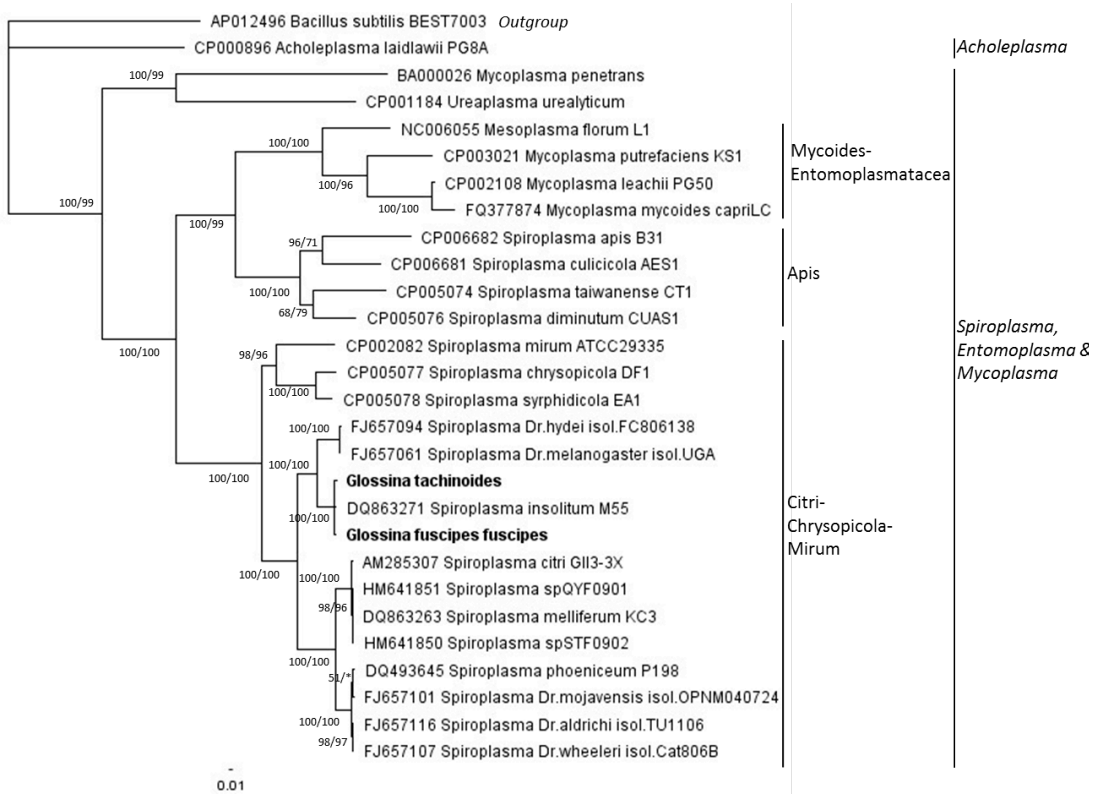


Figure 7.4 | Bayesian inference phylogeny of the *parE* gene.

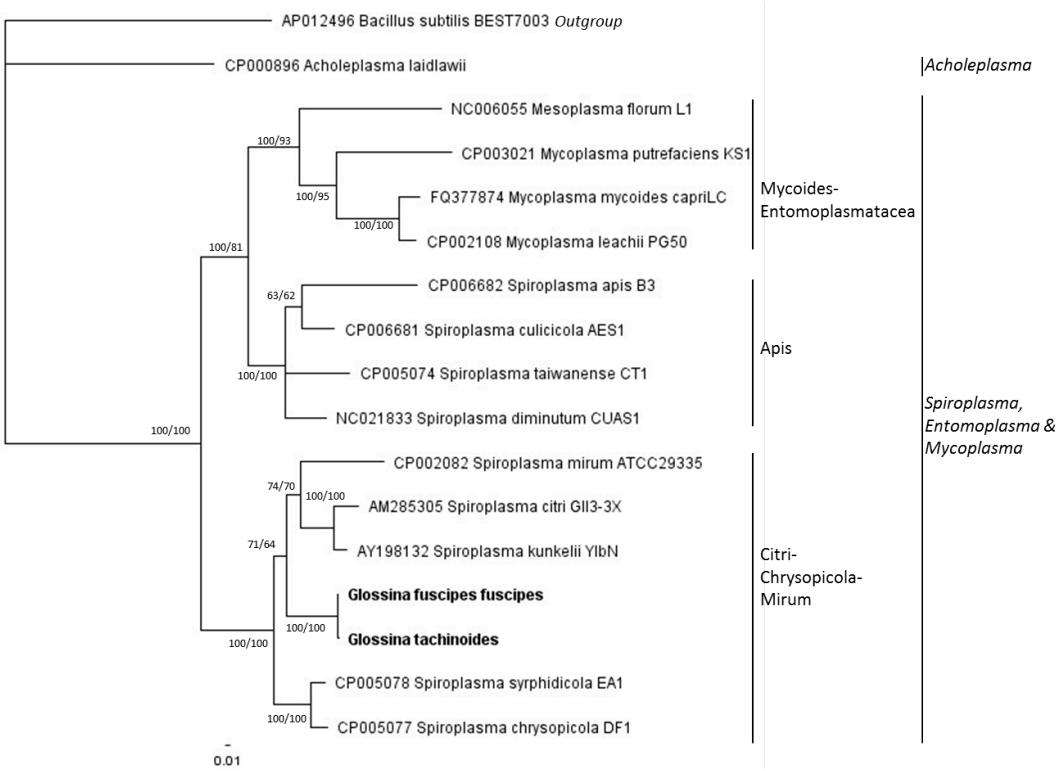


Figure 7.5 | Bayesian inference phylogeny of the *ftsZ* gene.

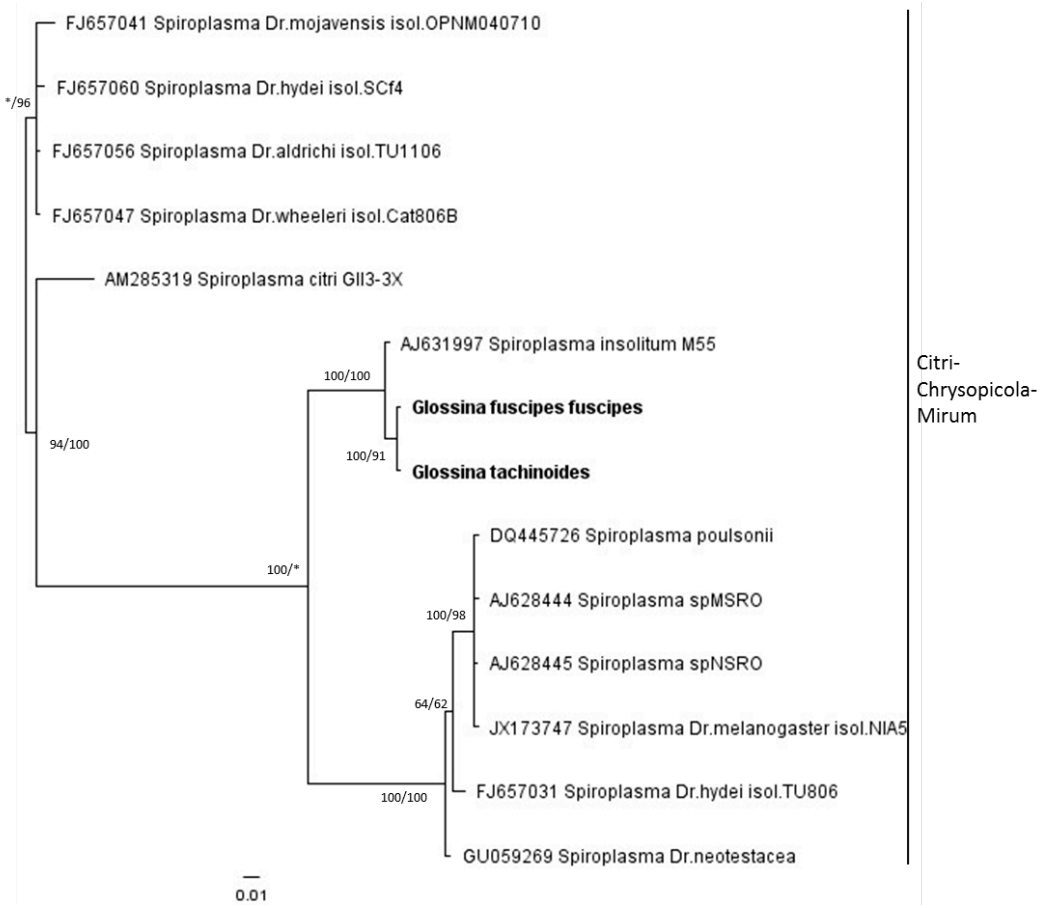


Figure 7.6 | Bayesian inference phylogeny of the *fruR* gene.

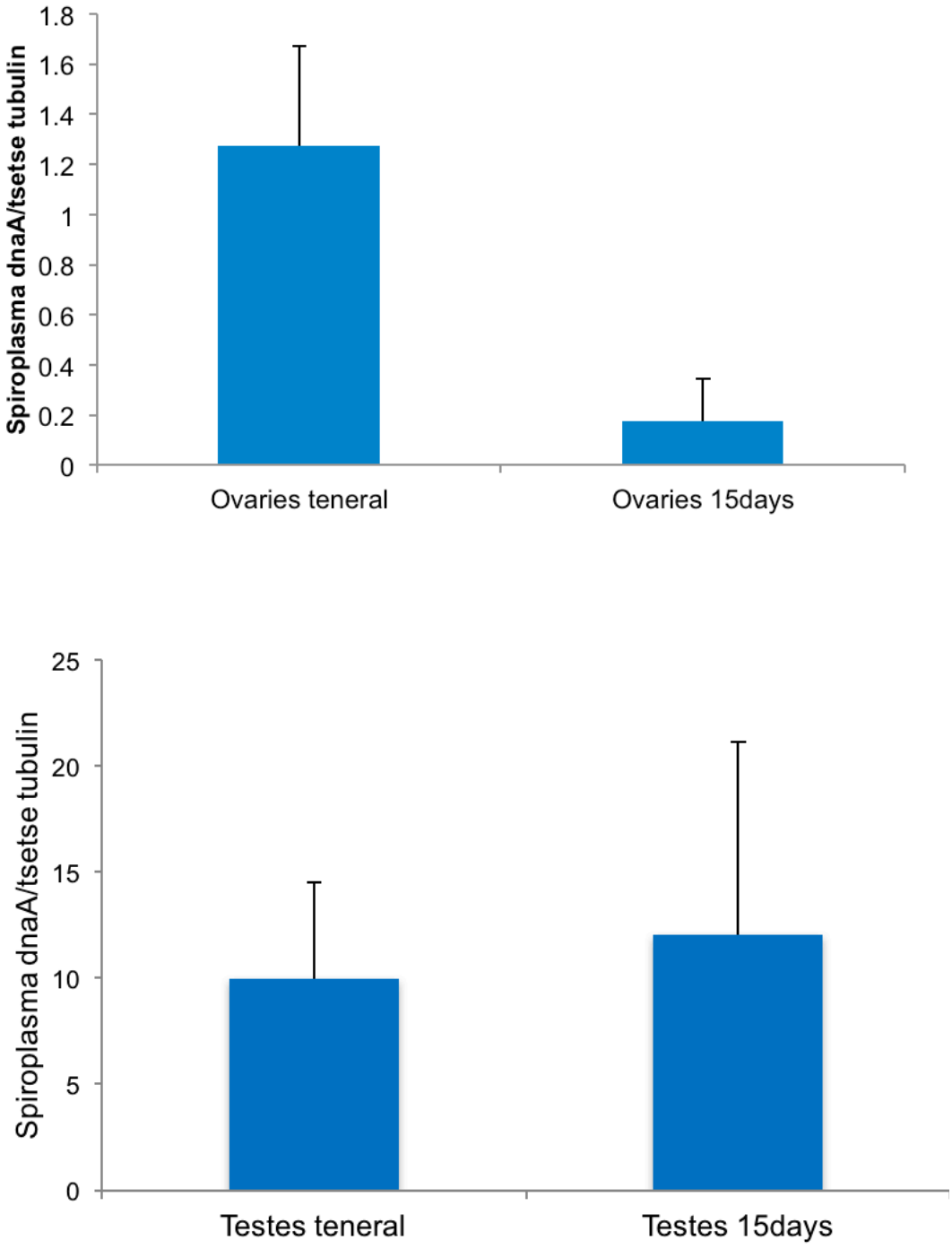


Figure 7.7 | *Spiroplasma dnaA* copy number normalized to the *Glossina β-tubulin* gene in **A** the ovaries of teneral and 15-day-old *G. f. fuscipes* females and **B** the testes of teneral and 15-day-old *G. f. fuscipes* males. Three replicates of five individuals were assessed per sex at each time point.

References

Abd-Alla AMM, Bergoin M, Parker AG, Maniania NK, Vlak JM, Bourtzis K, et al. (2013). Improving Sterile Insect Technique (SIT) for tsetse flies through research on their symbionts and pathogens. J Invertebr Pathol 112: S2–S10.

Aharon Y, Pasternak Z, Ben-Yosef M, Behar A, Lauzon C, Yuval B, et al. (2012). Phylogenetic, metabolic, and taxonomic diversities shape mediterranean fruit fly microbiotas during ontogeny. Appl Environ Microbiol 79: 303–313.

Akaike H. (1971). Autoregressive model fitting for control. Ann Inst Stat Math 23: 163–180.

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, et al. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. Nature Genetics 32: 402–407.

Aksoy E, Telleria EL, Echodu R, Wu Y, Okedi LM, Weiss BL, et al. (2014). Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. Appl Environ Microbiol 80: 4301–4312.

Aksoy S. (2011). Sleeping sickness elimination in sight: time to celebrate and reflect, but not relax. PLoS Negl Trop Dis 5: e1008.

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.

Alam U, Hyseni C, Symula RE, Brelsfoard C, Wu Y, Kruglov O, et al. (2012). Implications of microfauna-host interactions for trypanosome transmission dynamics in Glossina fuscipes fuscipes in Uganda. Appl Environ Microbiol 78: 4627–4637.

Alam U, Medlock J, Brelsfoard C, Pais R, Lohs C, Balmand S, et al. (2011). Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly Glossina morsitans. PLoS Pathog 7: e1002415.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. J Mol Biol 215: 403–410.

Ami EB, Yuval B, Jurkevitch E. (2009). Manipulation of the microbiota of mass-reared Mediterranean fruit flies Ceratitis capitata (Diptera: Tephritidae) improves sterile male sexual performance. ISME J 4: 28–37.

Amiot MJ, Fleuriet A, Macheix JJ. (1986). Importance and evolution of phenolic compounds in olive during growth and maturation. J Agric Food Chem 34: 823–826.

Anbutsu H, Fukatsu T. (2003). Population dynamics of male-killing and non-male-killing Spiroplasmas in Drosophila melanogaster. Appl Environ Microbiol 69: 1428–1434.

Anbutsu H, Fukatsu T. (2011). Spiroplasma as a model insect endosymbiont. Environ Microbiol Rep 3: 144–153.

Andersson SGE, Kurland CG. (1998). Reductive evolution of resident genomes. Trends Microbiol 6: 263–268.

Andongma AA, Wan L, Dong Y-C, li P, Desneux N, White JA, et al. (2015). Pyrosequencing reveals a shift in symbiotic bacteria populations across life stages of Bactrocera dorsalis. Sci Rep 5: 9470–6.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Ant T, Koukidou M, Rempoulakis P, Gong H-F, Economopoulos A, Vontas J, et al. (2012). Control of the olive fruit fly using genetics- enhanced sterile insect technique. BMC Biol 10: 51.

Appel HM, Martin MM. (1990). Gut redox conditions in herbivorous lepidopteran larvae. J Chem Ecol 16: 3277–3290.

Arambourg Y, Pralavorio R. (1970). Survie hivernale de Dacus oleae (Gmel). Ann Zool Ecol Anim 2: 659–662.

Ardia DR, Gantz JE, Schneider BC, Strebel S. (2012). Costs of immunity in insects: an induced immune response increases metabolic rate and decreases antimicrobial activity. Funct Ecol 26: 732–739.

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.

Augustinos AA, Kyritsis GA, Papadopoulos NT, Abd-Alla AMM, Cáceres C, Bourtzis K. (2015). Exploitation of the medfly gut microbiota for the enhancement of sterile insect technique: use of Enterobacter sp. in larval diet-based probiotic applications. PLoS ONE 10: e0136459.

Awmack CS, Leather SR. (2002). Host plant quality and fecundity in herbivorous insects. Annu Rev Entomol 47: 817–844.

Ayres JS. (2008). Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. Nat Rev Immunol 8: 889–895.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37: W202–8.

Ballard JWO, Melvin RG. (2007). Tetracycline treatment influences mitochondrial metabolism and mtDNA density two generations after treatment in Drosophila. Insect Mol Biol 16: 799–802.

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–S122.

Bateman AJ. (1948). Intra-sexual selection in Drosophila. Heredity 2: 349–368.

Bateman MA. (1972). The ecology of fruit flies. Annu Rev Entomol 17: 493–518.

Baumann P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol 59: 155–189.

Baumann P, Moran NA, Baumann L. (1997). The evolution and genetics of aphid endosymbionts. BioScience 47: 12–20.

Beard CB, Cordon-Rosales C, Durvasula RV. (2002). Bacterial symbionts of the triatominae and their potential use in control of Chagas disease transmission. Annu Rev Entomol 47: 123–141.

Behar A, Jurkevitch E, Yuval B. (2008). Bringing back the fruit into fruit fly–bacteria interactions. Mol Ecol 17: 1375–1386.

Behar A, Yuval B, Jurkevitch E. (2005). Enterobacteria-mediated nitrogen fixation in natural populations of the fruit fly Ceratitis capitata. Mol Ecol 14: 2637–2643.

Ben-Yosef M, Aharon Y, Jurkevitch E, Yuval B. (2010). Give us the tools and we will do the job: symbiotic bacteria affect olive fly fitness in a diet-dependent fashion. Proc Biol Sci 277: 1545–1552.

Ben-Yosef M, Pasternak Z, Jurkevitch E, Yuval B. (2014). Symbiotic bacteria enable olive flies (Bactrocera oleae) to exploit intractable sources of nitrogen. J Evol Biol 27: 2695–2705.

Ben-Yosef M, Pasternak Z, Jurkevitch E, Yuval B. (2015). Symbiotic bacteria enable olive fly larvae to overcome host defences. R Soc Open Sci 2: 150170–14.

Benemann JR. (1973). Nitrogen fixation in termites. Science 181: 164–165.

Bennett GM, Moran NA. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci USA 112: 10169–10176.

Bennett GM, Moran NA. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. Genome Biol Evol 5: 1675–1688.

Berasategui A, Shukla S, Salem H, Kaltenpoth M. (2016). Potential applications of insect symbionts in biotechnology. Appl Microbiol Biotechnol 100: 1567–1577.

Bi K, Huang H, Gu W, Wang J, Wang W. (2008). Phylogenetic analysis of Spiroplasmas from three freshwater crustaceans (Eriocheir sinensis, Procambarus clarkia and Penaeus vannamei) in China. J Invertebr Pathol 99: 57–65.

Biegel E, Schmidt S, González JM, Müller V. (2010). Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. Cell Mol Life Sci 68: 613–634.

Bolanos LM, Servin-Garciduenas LE, Martinez-Romero E. (2015). Arthropod-Spiroplasma relationship in the genomic era. FEMS Microbiol Ecol 91: 1–8.

Bourtzis K, Dobson SL, Xi Z, Rasgon JL, Calvitti M, Moreira LA, et al. (2014). Harnessing mosquito–Wolbachia symbiosis for vector and disease control. Acta Tropica 132: S150–S163.

Bourtzis K, Lees RS, Hendrichs J, Vreysen MJB. (2016). More than one rabbit out of the hat: Radiation, transgenic and symbiont-based approaches for sustainable management of mosquito and tsetse fly populations. Acta Tropica 157: 115–130.

Braendle C, Miura T, Bickel R, Shingleton AW, Kambhampati S, Stern DL. (2003). Developmental origin and evolution of bacteriocytes in the aphid– Buchnera symbiosis. PLoS Biol 1: e21.

Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U, et al. (2014). Presence of extensive Wolbachia symbiont insertions discovered in the genome of its host Glossina morsitans morsitans. PLoS Negl Trop Dis 8: e2728.

Bridges JR. (1981). Nitrogen-fixing bacteria associated with bark beetles. Microb Ecol 7: 131–137.

Broderick NA. (2016). Friend, foe or food? Recognition and the role of antimicrobial peptides in gut immunity and Drosophila-microbe interactions. Philos Trans R Soc Lond, B, Biol Sci 371: 20150295.

Broumas T, Haniotakis G, Liaropoulos C, Tomazou T, Ragoussis N. (2002). The efficacy of an improved form of the mass-trapping method, forthe control of the olive fruit fly, Bactrocera oleae (Gmelin) (Dipt., Tephritidae): pilot-scale feasibility studies. J Appl Entomol 126: 217–223.

Brownlie JC, Cass BN, Riegler M, Witsenburg JJ, Iturbe-Ormaetxe I, McGraw EA, et al. (2009). Evidence for metabolic provisioning by a common invertebrate endosymbiont, Wolbachia pipientis, during periods of nutritional stress. PLoS Pathog 5: e1000368.

Buchner P. (1965). *Endosymbiosis of animals with plant microorganisms*. John Wiley & Sons: New York, NY, USA.

Buchon N, Silverman N, Cherry S. (2014). Immunity in Drosophila melanogaster — from microbial recognition to whole-organism physiology. Nat Rev Immunol 14: 796–810.

Bursell E. (1967). The Excretion of Nitrogen in Insects. In: *Advances in Insect Physiology*. Elsevier: London, UK, pp 33–67.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26: 266–267.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7: 335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6: 1621–1624.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 108: 4516–4522.

Capuzzo C, Firrao G, Mazzon L, Squartini A, Girolami V. (2005). 'Candidatus Erwinia dacicola', a coevolved symbiotic bacterium of the olive fly Bactrocera oleae (Gmelin). Int J Syst Evol Microbiol 55: 1641–1647.

Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552.

Cayol J-P. (1999). Changes in sexual behavior and life history traits of tephritid species caused by mass-rearing processes. In: *Phylogeny and Evolution of Behavior*. Fruit Flies (Tephritidae). CRC Press, pp 843–860.

Chanbusarakum L, Ullman D. (2008). Characterization of bacterial symbionts in Frankliniella occidentalis (Pergande), Western flower thrips. J Invertebr Pathol 99: 318–325.

Chandler JA, James PM, Jospin G, Lang JM. (2014). The bacterial communities of Drosophila suzukii collected from undamaged cherries. PeerJ 2: e474–10.

Chandler JA, Morgan Lang J, Bhatnagar S, Eisen JA, Kopp A. (2011). Bacterial communities of diverse Drosophila species: ecological context of a host–microbe model system. PLoS Genet 7: e1002272.

Charlat S, Engelstädter J, Dyson EA, Hornett EA, Duplouy A, Tortosa P, et al. (2006). Competing selfish genetic elements in the butterfly Hypolimnas bolina. Current Biology 16: 2453–2458.

Charles H, Balmand S, Lamelas A, Cottret L, Pérez-Brocal V, Burdin B, et al. (2011). A genomic reappraisal of symbiotic function in the aphid/Buchnera symbiosis: reduced transporter sets and variable membrane organisations. PLoS ONE 6: e29096.

Chen B, Teh B-S, Sun C, Hu S, Lu X, Boland W, et al. (2016). Biodiversity and activity of the gut microbiota across the life history of the insect herbivore Spodoptera littoralis. Sci Rep 6: 29505.

Chen H, Boutros PC. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12: 1.

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.

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.

Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, O'neill SL, et al. (2000). Tissue distribution and prevalence of Wolbachia infections in tsetse flies, Glossina spp. Med Vet Entomol 14: 44–50.

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10: 563–569.

Chippindale AK, Leroi AM, Kim SB, Rose MR. (1993). Phenotypic plasticity and selection in Drosophila life-history evolution. I. Nutrition and the cost of reproduction. J Evol Biol 6: 171–193.

Chopra I, Roberts M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev 65: 232–260.

Christenson LD, Foote RH. (1960). Biology of fruit flies. Annu Rev Entomol 5: 171–192.

Chrostek E, Marialva MSP, Esteves SS, Weinert LA, Martinez J, Jiggins FM, et al. (2013). Wolbachia variants induce differential protection to viruses in Drosophila melanogaster: a phenotypic and phylogenomic analysis. PLoS Genet 9: e1003896.

Chrostek E, Marialva MSP, Yamada R, O'Neill SL, Teixeira L. (2014). High anti-viral protection without immune upregulation after interspecies Wolbachia transfer. PLoS ONE 9: e99025.

Chrostek E, Teixeira L. (2015). Mutualism breakdown by amplification of Wolbachia genes. PLoS Biol 13: e1002065.

Clark TB, Whitcomb RF, Tully JG, Mouches C, Saillard C, Bove JM, et al. (1985). Spiroplasma melliferum, a new species from the honeybee (Apis mellifera). Int J Syst Bacteriol 35: 296–308.

Clausen, Adeyemi, Bauer, Breloeer, Salchow, Staak. (1998). Host preferences of tsetse (Diptera: Glossinidae) based on bloodmeal identifications. Med Vet Entomol 12: 169–180.

Cotter PD, Hill C. (2003). Surviving the acid test: responses of gram-positive bacteria to low pH. Microbiol Mol Biol Rev 67: 429–53.

Currie CR, Scott JA, Summerbell RC, Malloch D. (2003). Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. Nature 423: 461–461.

D'Orazio SE, Collins CM. (1993). The plasmid-encoded urease gene cluster of the family Enterobacteriaceae is positively regulated by UreR, a member of the AraC family of transcriptional activators. J Bacteriol 175: 3459–3467.

Daane KM, Johnson MW. (2010). Olive fruit fly: managing an ancient pest in modern times. Annu Rev Entomol 55: 151–169.

Daane KM, Wang X, Nieto DJ, Pickett CH, Hoelmer KA, Blanchet A, et al. (2015). Classic biological control of olive fruit fly in California, USA: release and recovery of introduced parasitoids. BioControl 60: 317–330.

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.

Dantzer R. (2004). Cytokine-induced sickness behaviour: a neuroimmune response to activation of innate immunity. Eur J Pharmacol 500: 399–411.

Dantzer R, Kelley KW. (2007). Twenty years of research on cytokine-induced sickness behavior. Brain Behav Immun 21: 153–160.

De Vooght L, Caljon G, De Ridder K, Van Den Abbeele J. (2014). Delivery of a functional anti-trypanosome Nanobody in different tsetse fly tissues via a bacterial symbiont, Sodalis glossinidius. Microb Cell Fact 13: 156.

De Vooght L, Caljon G, Van Hees J, Van Den Abbeele J. (2015). Paternal transmission of a secondary symbiont during mating in the viviparous tsetse Fly. Mol Biol Evol 32: 1977–1980.

de Vries EJ, Jacobs G, Sabelis MW, Menken SBJ, Breeuwer JAJ. (2004). Diet-dependent effects of gut bacteria on their insect host: the symbiosis of Erwinia sp. and western flower thrips. Proc R Soc Lond B 271: 2171–2178.

Dennis JW, Durkin SM, Downie JEH, Hamill LC, Anderson NE, MacLeod ET. (2014). Sodalis glossinidius prevalence and trypanosome presence in tsetse from Luambe National Park, Zambia. Parasites & Vectors 7: 1–11.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–5072.

Dillon RJ, Dillon VM. (2004). The gut bacteria of insects: nonpathogenic interactions. Annu Rev Entomol 49: 71–92.

Ding Z, Du J, Ou J, Li W, Wu T, Xiu Y, et al. (2012). Classification of circulating hemocytes from the red swamp crayfish Procambarus clarkii and their susceptibility to the novel pathogen Spiroplasma eriocheiris in vitro. Aquaculture 356-357: 371–380.

Dohlen von CD, Kohler S, Alsop ST, McManus WR. (2001). Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. Nature 412: 433–436.

Donaldson GP, Lee SM, Mazmanian SK. (2015). Gut biogeography of the bacterial microbiota. Nat Rev Microbiol 14: 1–13.

Doudoumis V, Alam U, Aksoy E, Abd-Alla AMM, Tsiamis G, Brelsfoard C, et al. (2013). Tsetse-Wolbachia symbiosis: comes of age and has great potential for pest and disease control. J Invertebr Pathol 112: S94–S103.

Doudoumis V, Tsiamis G, Wamwiri F, Brelsfoard C, Alam U, Aksoy E, et al. (2012). Detection and characterization of Wolbachia infections in laboratory and natural populations of different species of tsetse flies (genus Glossina). BMC Microbiology 12: S3.

Douglas AE. (2011). Lessons from studying insect symbioses. Cell Host Microbe 10: 359–367.

Douglas AE. (2006). Phloem-sap feeding by animals: problems and solutions. J Exp Bot 57: 747–754.

Douglas AE. (2007). Symbiotic microorganisms: untapped resources for insect pest control. Trends Biotechnol 25: 338–342.

Douglas AE. (2009). The microbial dimension in insect nutritional ecology. Funct Ecol 23: 38–47.

Douglas AE. (2010). *The Symbiotic Habit*. Princeton University Press: Princeton, NJ, USA.

Drew R, Yuval B. (2000). The evolution of fruit fly feeding behavior. In: *Fruit Flies (Tephritidae) Phylogeny and Evolution of* Behavior. CRC Press: Boca Raton, FL, USA, pp 731–749.

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, et al. (2008). The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. BMC Biol 6: 27.

Durvasula RV, Gumbs A, Panackal A, Kruglov O, Aksoy S, Merrifield RB, et al. (1997). Prevention of insect-borne disease: An approach using transgenic symbiotic bacteria. Proc Natl Acad Sci USA 94: 3274–3278.

Dyer N, Lawton S, Ravel S, Choi K, Lehane M, Robinson A, et al. (2008). Molecular phylogenetics of tsetse flies (Diptera: Glossinidae) based on mitochondrial (COI, 16S, ND2) and nuclear ribosomal DNA sequences, with an emphasis on the palpalis group. Mol Phylogenet Evol 49: 227–239.

Economopoulos AP. (2002). The olive fruit fly, Bactrocera (Dacus) oleae (Gmelin) (Diptera: Tephritidae): its importance and control; previous SIT research and pilot testing. *Report to International Atomic Energy Agency (IAEA), Vienna, Austria* (2002): 44.

Economopoulos AP, Avtzis N, Zervas G, Tsitsipis J, Haniotakis G, Tsiropoulos G, et al. (1977). Experiments on the control of the olive fly, Dacus oleae (Gmel.), by the combined effect of insecticides and releases of gamma-ray sterilized insects. J Appl Entomol 83: 201–215.

Economopoulos AP, Haniotakis GE, Michelakis S, Tsiropoulos GJ, Zervas GA, Tsitsipis JA, et al. (1982). Population studies on the olive fruit fly, Dacus oleae (Gmel.) (Dipt., Tephritidae) in Western Crete. J Appl Entomol 93: 463–476.

Economopoulos AP, Zervas GA. (1982). The quality problem in olive flies produced for SIT experiments. *Report to International Atomic Energy Agency (IAEA), Vienna, Austria*: 357–368.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194–2200.

Engel P, Martinson VG, Moran NA. (2012). Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci USA 109: 11002–11007.

Engel P, Moran NA. (2013). The gut microbiota of insects – diversity in structure and function. FEMS Microbiol Rev 37: 699–735.

Estes AM. (2009). Life in a fly: the ecology and evolution of the olive fly endosymbiont, Candidatus Erwinia dacicola. PhD dissertation.

Estes AM, Hearn DJ, Bronstein JL, Pierson EA. (2009). The olive fly endosymbiont, 'Candidatus Erwinia dacicola,' switches from an intracellular existence to an extracellular existence during host insect development. Appl Environ Microbiol 75: 7097–7106.

Estes AM, Hearn DJ, Burrack HJ, Rempoulakis P, Pierson EA. (2012a). Prevalence of Candidatus Erwinia dacicola in wild and laboratory olive fruit fly populations and across developmental stages. Environ Entomol 41: 265–274.

Estes AM, Nestel D, Belcari A, Jessup A, Rempoulakis P, Economopoulos AP. (2012b). A basis for the renewal of sterile insect technique for the olive fly, Bactrocera oleae (Rossi). J Appl Entomol 136: 1–16.

Estes AM, Segura DF, Jessup A, Wornoayporn V, Pierson EA. (2014). Effect of the symbiont Candidatus Erwinia dacicola on mating success of the olive fly Bactrocera oleae (Diptera: Tephritidae). Int J Trop Insect Sci 34: S123–S131.

Facey PD, Méric G, Hitchings MD, Pachebat JA, Hegarty MJ, Chen X, et al. (2015). Draft genomes, phylogenetic reconstruction, and comparative genomics of two novel cohabiting bacterial symbionts isolated from Frankliniella occidentalis. Genome Biol Evol 7: 2188–2202.

Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E. (2002). Endosymbiotic bacteriaGroEL buffers against deleterious mutations. Nature 417: 398–398.

Farikou O, Njiokou F, Mbida Mbida JA, Njitchouang GR, Djeunga HN, Asonganyi T, 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: 115–121.

Farikou O, Thevenon S, Njiokou F, Allal F, Cuny G, Geiger A. (2011). Genetic diversity and population structure of the secondary symbiont of tsetse flies, Sodalis glossinidius, in sleeping sickness foci in Cameroon. PLoS Negl Trop Dis 5: e1281.

Feldhaar H. (2011). Bacterial symbionts as mediators of ecologically important traits of insect hosts. Ecol Entomol 36: 533–543.

Fletcher B. (1987). The biology of dacine fruit flies. Annu Rev Entomol 32: 115–144.

Fletcher BS. (1989). Life history strategies of tephritid fruit flies. In: *Fruit flies their biology, natural enemies, and control*. University of Amsterdam: Amsterdam, Holland, pp 105–118.

Fletcher BS, Pappas S, Kapatos E. (1978). Changes in the ovaries of olive flies (Dacus oleae (Gmelin)) during the summer, and their relationship to temperature, humidity and fruit availability. Ecol Entomol 3: 99–107.

Frentiu FD, Zakir T, Walker T, Popovici J, Pyke AT, van den Hurk A, et al. (2014). Limited dengue virus replication in field-collected Aedes aegypti mosquitoes infected with Wolbachia. PLoS Negl Trop Dis 8: e2688.

Fukatsu T, Ishikawa H. (1992). A novel eukaryotic extracellular symbiont in an aphid, Astegopteryx styraci (Homoptera, Aphididae, Hormaphidinae). J Insect Physiol 38: 765–773.

Fukatsu T, Nikoh N. (1998). Two intracellular symbiotic bacteria from the Mulberry Psyllid *Anomoneura mori* (Insecta, Homoptera). *Appl Environ Microbiol* **64**: 3599–3606.

Fukatsu T, Nikoh N. (2000). Endosymbiotic microbiota of the bamboo Pseudococcid *Antonina crawii* (Insecta, Homoptera). *Appl Environ Microbiol* **66**: 643–650.

Fytizas E, Tzanakakis ME. (1966). Some effects of streptomycin, when added to the adult food, on the adults of Dacus oleae (Diptera: Tephritidae) and their progeny. Ann Entomol Soc Am 59: 269–273.

Garcia JR, Gerardo NM. (2014). The symbiont side of symbiosis: do microbes really benefit? Front Microbiol 5: 510.

Gasparich GE. (2010). Spiroplasmas and phytoplasmas: microbes associated with plant hosts. Biologicals 38: 193–203.

Geiger A, Cuny G, Frutos R. (2005). Two Tsetse fly species, Glossina palpalis gambiensis and Glossina morsitans morsitans, carry genetically distinct populations of the secondary symbiont Sodalis glossinidius. Appl Environ Microbiol 71: 8941–8943.

Geiger A, Fardeau M-L, Grebaut P, Vatunga G, Josénando T, Herder S, 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: 1364–1370.

Geiger A, Fardeau M-L, Njiokou F, Joseph M, Asonganyi T, Ollivier B, et al. (2011). Bacterial diversity associated with populations of Glossina spp. from Cameroon and distribution within the Campo sleeping sickness focus. Microb Ecol 62: 632–643.

Geiger A, Fardeau M-L, Njiokou F, Ollivier B. (2013). Glossina spp. gut bacterial flora and their putative role in fly-hosted trypanosome development. Front Cell Infect Microbiol 3: 34.

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.

Geiger A, Ravel S, Mateille T, Janelle J, Patrel D, Cuny G, et al. (2006). Vector competence of Glossina palpalis gambiensis for Trypanosoma brucei s.l. and genetic diversity of the symbiont Sodalis glossinidius. Mol Biol Evol 24: 102–109.

Geiger A, Tchicaya B, Rihet P. (2015). Technical data of the transcriptomic analysis performed on tsetse fly symbionts, Sodalis glossinidius and Wigglesworthia glossinidia, harbored, respectively by non-infected, Trypanosoma brucei gambiense infected and self-cured Glossina palpalis gambiensis tsetse flies. Genomics Data 4: 133–136.

Genc H, Nation JL. (2008). Maintaining Bactrocera oleae (Gmelin.) (Diptera: Tephritidae) colony on its natural host in the laboratory. J Pest Sci 81: 167–174.

Gibson CM, Hunter MS. (2010). Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. Ecol Letters 13: 223–234.

Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A, et al. (2003). The genome sequence of Blochmannia floridanus: comparative analysis of reduced genomes. Proc Natl Acad Sci USA 100: 9388–9393.

Goto S, Anbutsu H, Fukatsu T. (2006). Asymmetrical Interactions between Wolbachia and Spiroplasma Endosymbionts Coexisting in the Same Insect Host. Appl Environ Microbiol 72: 4805–4810.

Gotoh T, Noda H, Ito S. (2006). Cardinium symbionts cause cytoplasmic incompatibility in spider mites. *Heredity* **98**: 13–20

Greenberg B, Klowden M. (1972). Enteric bacterial interactions in insects. Am J Clin Nutr 25: 1459–1466.

Griessl MH, Schmid B, Kassler K, Braunsmann C, Ritter R, Barlag B, et al. (2013). Structural insight into the giant Ca2+-binding adhesin SiiE: implications for the adhesion of Salmonella enterica to polarized epithelial cells. Structure 21: 741–752.

Grimont PAD, Grimont F. (1978). The genus Serratia. Annu Rev Microbiol 32: 221–248.

Gross R, Vavre F, Heddi A, Hurst GDD, Zchori-Fein E, Bourtzis K. (2009). Immunity and symbiosis. Mol Microbiol 73: 751–759.

Guz N, Attardo GM, Wu Y, Aksoy S. (2007). Molecular aspects of transferrin expression in the tsetse fly (Glossina morsitans morsitans). J Insect Physiol 53: 715–723.

Ha EM, Oh CT, Bae YS, Lee WJ. (2005). A direct role for dual xxidase in Drosophila gut immunity. Science 310: 847–850.

Hagen KS. (1966). Dependence of the olive fly, Dacus oleae, larvae on symbiosis with Pseudomonas savastanoi for the utilization of olive. Nature 209: 423–424.

Hagen KS, Santas L, Tsecouras A. (1963). A technique of culturing the olive fly Dacus oleae Gmel. on synthetic media under xenic conditions. IAEA: Athens, Greece.

Hammer TJ, McMillan WO, Fierer N. (2014). Metamorphosis of a butterfly-associated bacterial community. PLoS ONE 9: e86995.

Haniotakis GE. (2005). Olive pest control: present status and prospects. Int Prot Olive Crops IOBC WPRS Bull 1–9.

Hanner R, Fugate M. (1997). Branchiopod phylogenetic reconstruction from 12S rDNA sequence data. *J Crust Biol* 17: 174–183.

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

Hart BL. (1988). Biological basis of the behavior of sick animals. Neurosci Biobehav Rev 12: 123–137.

Hartmann M, Howes CG, Veldre V, Schneider S, Vaishampayan PA, Yannarell AC, et al. (2011). V-REVCOMP: automated high-throughput detection of reverse complementary 16S rRNA gene sequences in large environmental and taxonomic datasets. FEMS Microbiol Lett 319: 140–145.

Harumoto T, Anbutsu H, Fukatsu T. (2014). Male-killing Spiroplasma induces sex-specific cell death via host apoptotic pathway. *PLOS Pathog* 10: e1003956.

Haselkorn TS, Cockburn SN, Hamilton PT, Perlman SJ, Jaenike J. (2013). Infectious adaptation: potential host range of a defensive endosymbiont in Drosophila. Evolution 67: 934–945.

Haselkorn TS, Markow TA, Moran NA. (2009). Multiple introductions of the Spiroplasmabacterial endosymbiont into Drosophila. Mol Ecol 18: 1294–1305.

Hautefort I, Proenca MJ, Hinton JCD. (2003). Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during infection of mammalian cells. Appl Environ Microbiol 69: 7480–7491.

Hayashi M, Watanabe M, Yukuhiro F, Nomura M, Kageyama D. (2016). A nightmare for males? A maternally transmitted male-killing bacterium and strong female bias in a green lacewing population. PLoS ONE 11: e0155794.

Hedges LM, Brownlie JC, O'neill SL, Johnson KN. (2008). Wolbachia and virus protection in insects. Science 322: 702–702.

Hedges LM, Yamada R, O'Neill SL, Johnson KN. (2012). The small interfering RNA pathway is not essential for Wolbachia-mediated antiviral protection in Drosophila melanogaster. Appl Environ Microbiol 78: 6773–6776.

Heres A, Lightner DV. (2010). Phylogenetic analysis of the pathogenic bacteria Spiroplasma penaei based on multilocus sequence analysis. J Invertebr Pathol 103: 30–35.

Herren JK, Lemaitre B. (2011). Spiroplasma and host immunity: activation of humoral immune responses increases endosymbiont load and susceptibility to certain Gram-negative bacterial pathogens in Drosophila melanogaster. Cell Microbiol 13: 1385–1396.

Herren JK, Paredes JC, Schüpfer F, Arafah K, Bulet P, Lemaitre B. (2014). Insect endosymbiont proliferation is limited by lipid availability. eLife 3: 4570.

Hoek TA, Axelrod K, Biancalani T, Yurtsev EA, Liu J, Gore J. (2016). Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. PLoS Biol 14: e1002540.

Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, Muzzi F, et al. (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. Nature 476: 454–457.

Hongoh Y, Ishikawa H. (1997). Uric acid as a nitrogen resource for the brown planthopper, Nilaparvata lugens: studies with synthetic diets and aposymbiotic insects. Zool Sci 14: 581–586.

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. (2006). Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol 4: e337.

Howick VM, Lazzaro BP. (2014). Genotype and diet shape resistance and tolerance across distinct phases of bacterial infection. BMC Evol Biol 14: 56.

Hughes GL, Dodson BL, Johnson RM, Murdock CC, Tsujimoto H, Suzuki Y, et al. (2014). Native microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes. Proc Natl Acad Sci USA 111: 12498–12503.

Humphery-Smith I, Grulet O, Le Goff F, Chastel C. (1991). Spiroplasma (Mollicutes: Spiroplasmataceae) pathogenic for Aedes aegypti and Anopheles stephensi (Diptera: Culicidae). J Med Entomol 28: 219–222.

Hurst G. (2000). Male-killing bacteria in insects: mechanisms, incidence, and implications. Emerg Infect Dis 6: 329–336.

Hurst GDD, Anbutsu H, Kutsukake M, Fukatsu T. (2003). Hidden from the host: Spiroplasma bacteria infecting Drosophila do not cause an immune response, but are suppressed by ectopic immune activation. Insect Mol Biol 12: 93–97.

Husnik F, McCutcheon JP. (2016). Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. Proc Natl Acad Sci USA 201603910.

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153: 1567–1578.

Husník F, Chrudimský T, Hypša V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. BMC Biol 9: 87.

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

International Glossina Genome Initiative. (2014). Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis. Science 344: 380–386.

Ishida H, Murai T, Sonoda S, Yoshida H, Izumi Y, Tsumuki H. (2003). Effects of temperature and photoperiod on development and oviposition of Frankliniella occidentalis (Pergande)(Thysanoptera: Thripidae). Appl Entomol Zool 38: 65–68.

Jaenike J. (2009). Coupled population dynamics of endosymbionts within and between hosts. Oikos 118: 353–362.

Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. (2010). Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. Science 329: 212–215.

Janson EM, Stireman JO, Singer MS, Abbot P. (2008). Phytophagous insect–microbe mutualisms and adaptive evolutionary diversification. Evolution, 62: 997-1012.

Johnston PR, Rolff J. (2015). Host and symbiont jointly control gut microbiota during complete metamorphosis. PLoS Pathog 11: e1005246.

Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle.

Kaiwa N, Hosokawa T, Nikoh N, Tanahashi M, Moriyama M, Meng X-Y, et al. (2014). Symbiont-supplemented maternal investment underpinning host's ecological adaptation. Current Biology 24: 2465–2470.

Kantheti P, Jayarama KS, Chandra HS. (1996). Developmental analysis of a female-specific 16s rRNA gene from mycetome-associated endosymbionts of a mealybug, Planococcus lilacinus. Insect Biochemistry and Molecular Biology 26: 997–1009.

Kapatos ET, Fletcher BS. (1984). The phenology of the olive fly, Dacus oleae (Gmel.) (Diptera, Tephritidae), in Corfu. J Appl Entomol 97: 360–370.

Kashima T, Nakamura T, Tojo S. (2006). Uric acid recycling in the shield bug, Parastrachia japonensis (Hemiptera: Parastrachiidae), during diapause. J Insect Physiol 52: 816–825.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Kembel SW, Wu M, Eisen JA, Green JL. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol 8:e1002743.

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.

Kikuchi Y, Hosokawa T, Nikoh N, Meng X-Y, Kamagata Y, Fukatsu T. (2009). Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. BMC Biol 7: 2.

Kim JK, Kwon JY, Kim SK, Han SH, Won YJ, Lee JH, et al. (2014). Purine Biosynthesis, Biofilm Formation, and Persistence of an Insect-Microbe Gut Symbiosis. Appl Environ Microbiol 80: 4374–4382.

Kim JK, Lee JB, Huh YR, Jang HA, Kim C-H, Yoo JW, et al. (2015). Burkholderia gut symbionts enhance the innate immunity of host Riptortus pedestris. Dev Comp Immunol 53: 265–269.

Klassen W, Curtis CF. (2005). History of the Sterile Insect Technique. In: *Sterile Insect Technique*. Springer-Verlag: Berlin, Germany, pp 39–68.

Koga R, Bennett GM, Cryan JR, Moran NA. (2013). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. Environ Microbiol 15: 2073–2081.

Koga R, Meng XY, Tsuchida T, Fukatsu T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. Proc Natl Acad Sci USA 109: E1230–E1237.

Koga R, Moran NA. (2014). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. ISME J 1–10.

Kono M, Koga R, Shimada M, Fukatsu T. (2008). Infection dynamics of co-existing Beta- and Gammaproteobacteria in the nested endosymbiotic system of mealybugs. Appl Environ Microbiol 74: 4175–4184.

Koo M-S, Lee J-H, Rah S-Y, Yeo W-S, Lee J-W, Lee K-L, et al. (2003). A reducing system of the superoxide sensor SoxR in Escherichia coli. The EMBO Journal 22: 2614–2622.

Kounatidis I, Crotti E, Sapountzis P, Sacchi L, Rizzi A, Chouaia B, et al. (2009). Acetobacter tropicalis Is a major symbiont of the olive fruit fly (Bactrocera oleae). Appl Environ Microbiol 75: 3281–3288.

Koveos DS, Tzanakakis ME. (1993). Diapause aversion in the adult olive fruit fly through effects of the host fruit, bacteria, and adult diet. Ann Entomol Soc Am 86: 668–673.

Koveos DS, Tzanakakis ME. (1990). Effect of the presence of olive fruit on ovarian maturation in the olive fruit fly, Dacus oleae, under laboratory conditions. Entomol Exp Appl 55: 161–168.

Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. (2013). Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front Genet 4: 237.

Kutzer MAM, Armitage SAO. (2016). The effect of diet and time after bacterial infection on fecundity, resistance, and tolerance in Drosophila melanogaster. Ecol Evol 6: 4229–4242.

Kwong WK, Engel P, Koch H, Moran NA. (2014). Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc Natl Acad Sci USA 111: 11509–11514.

Kwong WK, Moran NA. (2015). Evolution of host specialization in gut microbes: the bee gut as a model. Gut Microbes 6: 214–220.

Lambert JD, Moran NA. (1998). Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. Proc Natl Acad Sci USA 95: 4458–4462.

Lambrou PD, Tzanakakis ME. (1978). Inhibition of larval growth of Dacus oleae (Diptera: Tephritidae) by streptomycin. Effect of treating the parents. Entomol Exp Appl 23: 163–170.

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

Larkum AWD, Lockhart PJ, Howe CJ. (2007). Shopping for plastids. Trends Plant Sci 12: 189–195.

Latasa C, Roux A, Toledo-Arana A, Ghigo J-M, Gamazo C, Penadés JR, et al. (2005). BapA, a large secreted protein required for biofilm formation and host colonization of Salmonella enterica serovar Enteritidis. Mol Microbiol 58: 1322–1339.

Lauzon C. (2003). Symbiotic relationships of Tephritids. In: *Insect Symbiosis*. CRC Press, Boca Raton, FL, USA, pp 115–128.

Lauzon CR, Potter SE. (2012). Description of the irradiated and nonirradiated midgut of Ceratitis capitata Wiedemann (Diptera: Tephritidae) and Anastrepha ludens Loew (Diptera: Tephritidae) used for sterile insect technique. J Pest Sci 85: 217–226.

Leak SG. (1999). *Tsetse biology and ecology: their role in the epidemiology and control of trypanosomosis*. CABI publishing: Wallingford, UK.

Lee FJ, Rusch DB, Stewart FJ, Mattila HR, Newton ILG. (2015). Saccharide breakdown and fermentation by the honey bee gut microbiome. Environ Microbiol 17: 796–815.

Lefevre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A. (2004). Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial Replacement. Mol Biol Evol 21: 965–973.

Lemaitre B, Hoffmann J. (2007). The host defense of Drosophila melanogaster. Annu Rev Immunol 25: 697–743.

Li L, CJ S, Roos DS. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13: 2178–2189.

Liang T, Li X, Du J, Yao W, Sun G, Dong X, et al. (2011). Identification and isolation of a spiroplasma pathogen from diseased freshwater prawns, Macrobrachium rosenbergii, in China: A new freshwater crustacean host. Aquaculture 318: 1–6.

Lin DY-W, Huang S, Chen J. (2015). Crystal structures of a polypeptide processing and secretion transporter. Nature 523: 425–430.

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.

Lizé A, McKay R, Lewis Z. (2012). Gut microbiota and kin recognition. Trends Ecol Evol 1–2.

Lloyd AC, Drew R, Teakle DS, Hayward AC. (1986). Bacteria associated with some Dacus species (Diptera: Tephritidae) and their host fruit in Queensland. Aust J Biol Sci 39: 361–368.

Lukasik P, Guo H, van Asch M, Ferrari J, Godfray HCJ. (2013). Protection against a fungal pathogen conferred by the aphid facultative endosymbionts Rickettsia and Spiroplasma is expressed in multiple host genotypes and species and is not influenced by co-infection with another symbiont. J Evol Biol 26: 2654–2661.

Margaritopoulos JT, Skavdis G, Kalogiannis N, Nikou D, Morou E, Skouras PJ, et al. (2008). Efficacy of the pyrethroid alpha-cypermethrin against Bactrocera oleae populations from Greece, and improved diagnostic for an iAChE mutation. Pest Manag Sci 64: 900–908.

Martin JP, Beyerlein A, Dacks AM, Reisenman CE, Riffell JA, Lei H, et al. (2011). The neurobiology of insect olfaction: sensory processing in a comparative context. Prog Neurobiol 95: 427–447.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17: 10.

Martinez J, Longdon B, Bauer S, Chan Y-S, Miller WJ, Bourtzis K, et al. (2014). Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of Wolbachia strains. PLoS Pathog 10: e1004369.

Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA. (2010). A simple and distinctive microbiota associated with honey bees and bumble bees. Mol Ecol 20: 619–628.

Martinson VG, Moy J, Moran NA. (2012). Establishment of characteristic gut bacteria during development of the honeybee worker. Appl Environ Microbiol 78: 2830–2840.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. (2012). PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 2012 13:1 13: 1.

Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow TA, Moran NA. (2006). Heritable endosymbionts of *Drosophila*. *Genetics* 174: 363–376.

Matsuura Y, Hosokawa T, Serracin M, Tulgetske GM, Miller TA, Fukatsu T. (2014). Bacterial symbionts of a devastating coffee plant pest, the stinkbug *Antestiopsis thunbergii* (Hemiptera: Pentatomidae). Appl Environ Microbiol 80: 3769–3775.

Maudlin I. (1982). Inheritance of susceptibility to Trypanosoma congolense infection in Glossina morsitans. Ann Trop Med Parasitol 76: 225–227.

Maudlin I, Ellis DS. (1985). Association between intracellular rickettsial-like infections of midgut cells and susceptibility to Trypanosome infection in Glossina spp. Zeitschrift fur Parasitenkunde 71: 683–687.

McCutcheon JP, Dohlen von CD. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Current Biology 21: 1366–1372.

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

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

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci USA 110: 3229–3236.

McGraw EA, O'Neill SL. (2013). Beyond insecticides: new thinking on an ancient problem. Nat Rev Microbiol 11: 181–193.

McInnis DO, Lance DR, Jackson CG. (1996). Behavioral resistance to the Sterile Insect Technique by mediterranean fruit fly (Diptera: Tephritidae) in Hawaii. Ann Entomol Soc Am 89: 739–744.

McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, Foster KR. (2016). Host selection of microbiota via differential adhesion. Cell Host Microbe 19: 550–559.

McMurdie PJ, Holmes S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8: e61217.

Mediannikov O, Audoly G, Diatta G, Trape JF, Raoult D. (2012). New Rickettsia sp. in tsetse flies from Senegal. Comp Immunol Microbiol Infect Dis 35: 145–150.

Mira A, Ochman H, Moran NA. (2001). Deletional bias and the evolution of bacterial genomes. Trends in Genetics 17: 589–596.

Montenegro H, Solferini VN, Klaczko LB, Hurst GDD. (2005). Male-killing Spiroplasma naturally infecting Drosophila melanogaster. Insect Mol Biol 14: 281–287.

Moran NA. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci USA 93: 2873–2878.

Moran NA. (1994). Adaptation and constraint in the complex life cycles of animals. Ann Zool Ecol Anim 25: 573–600.

Moran NA. (2002). Microbial minimalism: genome reduction in bacterial pathogens. Cell 108: 583–586.

Moran NA. (2007). Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci USA 104 Suppl 1: 8627–8633.

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

Moran NA, McLaughlin HJ, Sorek R. (2009). The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323: 379–382.

Moran NA, Russell JA, Koga R, Fukatsu T. (2005). Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects. Appl Environ Microbiol 71: 3302–3310.

Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, et al. (2009). A Wolbachia symbiont in Aedes aegypti limits infection with Dengue, Chikungunya, and Plasmodium. Cell 139: 1268–1278.

Morgan E, Bowen AJ, Carnell SC, Wallis TS, Stevens MP. (2007). SiiE Is secreted by the Salmonella enterica serovar typhimurium pathogenicity island 4-encoded secretion system and contributes to intestinal colonization in cattle. Infect Immun 75: 1524–1533.

Mori M, Ponce-de-Leon M, Peretó J, Montero FM. (2016). Metabolic complementation in bacterial communities: necessary conditions and optimality. bioRxiv 064121.

Morris JJ, Lenski RE, Zinser ER. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio 3: e00036–12.

Morrow JL, Frommer M, Shearman DCA, Riegler M. (2015). The microbiome of field-caught and laboratory-adapted Australian tephritid fruit fly species with different host plant use and specialisation. Microb Ecol 70: 498–508.

Mouches C, Bove JM, Albisetti J. (1984). Pathogenicity of Spiroplasma apis and other spiroplasmas for honey-bees in southwestern France. Ann Microbiol (Paris) 135A: 151–155.

Mound LA, Teulon DAJ. (1995). Thysanoptera as Phytophagous Opportunists. In*: Thrips Biology and Management*. Springer US: Boston, MA, USA, pp 3–19.

Mulrooney SB, Hausinger RP. (1990). Sequence of the Klebsiella aerogenes urease genes and evidence for accessory proteins facilitating nickel incorporation. J Bacteriol 172: 5837–5843.

Murdock CC, Blanford S, Hughes GL, Rasgon JL, Thomas MB. (2014). Temperature alters Plasmodium blocking by Wolbachia. Sci Rep 4: 3932.

Mushegian AA, Ebert D. (2016). Rethinking ‘mutualism’ in diverse host-symbiont communities. Bioessays 38: 100–108.

Nai Y-S, Su P-Y, Hsu Y-H, Chiang C-H, Kim JS, Chen Y-W, et al. (2014). A new spiroplasma isolate from the field cricket (Gryllus bimaculatus) in Taiwan. J Invertebr Pathol 120: 4–8.

Nardi JB, Mackie RI, Dawson JO. (2002). Could microbial symbionts of arthropod guts contribute significantly to nitrogen fixation in terrestrial ecosystems? J Insect Physiol 48: 751–763.

Naum M, Brown EW, Mason-Gamer RJ. (2008). Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the Enterobacteriaceae? J Mol Evol 66: 630–642.

Neuenschwander P, Michelakis S, Bigler F. (1981). Abiotic factors affecting mortality of Dacus oleae larvae and pupae in the soil. Entomol Exp Appl 30: 1-9.

Neuenschwander P, Michelakis S, Kapatos ET. (1986). Tephritidae: Dacus oleae (Gmel). Entomologie oleicole 1: 115–159.

Nikolenko SI, Korobeynikov AI, Alekseyev MA. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics 14: 1.

Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. (2005). Bacterial genome size reduction by experimental evolution. Proc Natl Acad Sci USA 102: 12112–12116.

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

Nunan LM. (2005). Spiroplasma penaei sp. nov., associated with mortalities in Penaeus vannamei, Pacific white shrimp. Int J Syst Evol Microbiol 55: 2317–2322.

Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20: 714–737.

Nyholm SV, Graf J. (2012). Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. Nat Rev Microbiol 10: 815–827.

O'Toole G, Kaplan HB, Kolter R. (2000). Biofilm formation as microbial development. Annu Rev Microbiol 54: 49–79.

Okonechnikov K, Conesa A, García-Alcalde F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292–294.

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.

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

Oliver KM, Moran NA, Hunter MS. (2006). Costs and benefits of a superinfection of facultative symbionts in aphids. Proc R Soc Lond B 273: 1273–1280.

Oliver KM, Russell JA, Moran NA, Hunter MS. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc Natl Acad Sci USA 100: 1803–1807.

Otti O, Gantenbein-Ritter I, Jacot A, Brinkhof MWG. (2011). Immune response increases predation risk. Evolution 66: 732–739.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42: D206–14.

Paredes JC, Herren JK, Schüpfer F, Lemaitre B. (2016). The role of lipid competition for endosymbiont-mediated protection against parasitoid wasps in Drosophila. mBio 7: e01006–16.

Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, et al. (2015). Genome sequence of the Drosophila melanogaster male-killing Spiroplasma strain MSRO endosymbiont. mBio 6: e02437–14–12.

Parker BO, Marinus MG. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in Escherichia coli. Proc Natl Acad Sci USA 89: 1730–1734.

Patino-Navarrete R, Piulachs MD, Belles X, Moya A, Latorre A, Pereto J. (2014). The cockroach Blattella germanica obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. Biol Lett 10: 20140407.

Perlman SJ, Kelly SE, Hunter MS. (2008). Population biology of cytoplasmic incompatibility: maintenance and spread of Cardinium symbionts in a parasitic wasp. Genetics 178: 1003-1011.

Petri L. (1909). Ricerche sopra i batteri intestinali della mosca olearia. Memorie della Regia Stazione di Patologia Vegetale di Roma.

Pfeiffenberger C, Lear BC, Keegan KP, Allada R. (2010). Locomotor activity level monitoring using the Drosophila Activity Monitoring (DAM) System. Cold Spring Harbor Protocols, 11: pdb-prot5518.

Pinto-Tomas AA, Anderson MA, Suen G, Stevenson DM, Chu FST, Cleland WW, et al. (2009). Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science 326: 1120–1123.

Pontes MH, Dale C. (2006). Culture and manipulation of insect facultative symbionts. Trends Microbiol 14: 406–412.

Potrikus CJ, Breznak JA. (1981). Gut bacteria recycle uric acid nitrogen in termites: A strategy for nutrient conservation. Proc Natl Acad Sci USA 78: 4601–4605.

Potrikus CJ, Breznak JA. (1980). Uric acid in wood-eating termites. Insect Biochemistry 10: 19–27.

Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26: 1641–1650.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–6.

R Core Team. (2015). A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rakoff-Nahoum S, Foster KR, Comstock LE. (2016). The evolution of cooperation within the gut microbiota. Nature 533: 255–259.

Råberg L, Sim D, Read AF. (2007). Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. Science 318: 812–814.

Reinhardt C, Steiger R, Hecker H. (1972). Ultrastructural study of the midgut mycetome-bacteroids of the tsetse flies Glossina morsitans, G. Fuscipes, and G, brevipalpis (Diptera, Brachycera). Acta Tropica 29: 280–288.

Reis MD. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res 32: 5036–5044.

Reva O, Tümmler B. (2008). Think big--giant genes in bacteria. Environ Microbiol 10: 768–777.

Rincon-Enriquez G, Crété P, Barras F, Py B. (2008). Biogenesis of Fe/S proteins and pathogenicity: IscR plays a key role in allowing Erwinia chrysanthemi to adapt to hostile conditions. Mol Microbiol 67: 1257–1273.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature 499: 431–437.

Rolff J. (2002). Bateman's principle and immunity. Proc R Soc Lond B 269: 867–872.

Rolff J, Armitage SAO, Coltman DW. (2005). Genetic constraints and sexual dimorphism in immune defense. Evolution 59: 1844–1850.

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

Rossi P, Ricci I, Cappelli A, Damiani C, Ulissi U, Mancini MV, et al. (2015). Mutual exclusion of Asaia and Wolbachia in the reproductive organs of mosquito vectors. Parasites & Vectors 8: 278–10.

Roulston TH, Cane JH. (2000). Pollen nutritional content and digestibility for animals. Plant Syst and Evol 222: 187–209.

Russell JA, Moran NA. (2006). Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proc R Soc Lond B 273: 603–610.

Sabree ZL, Huang CY, Okusu A, Moran NA, Normark BB. (2013). The nutrient supplying capabilities of Uzinura, an endosymbiont of armoured scale insects. Environ Microbiol 15: 1988–1999.

Sabree ZL, Kambhampati S, Moran NA. (2009). Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont. Proc Natl Acad Sci USA 106: 19521–19526.

Sacantanis K. (1957). The scientific bases of a rational control of the olive fruit fly (Dacus oleae Gmel.). Geoponika 3: 219–225.

Sacchetti P, Ghiardi B, Granchietti A, Stefanini FM, Belcari A. (2013). Development of probiotic diets for the olive fly: evaluation of their effects on fly longevity and fecundity. Ann Appl Biol 164: 138–150.

Sacchetti P, Granchietti A, Landini S, Viti C, Giovannetti L, Belcari A. (2008). Relationships between the olive fly and bacteria. J Appl Entomol 132: 682–689.

Saglio P, Lhospital M, Lafleche D, Dupont G, Bove JM, Tully JG, et al. (1973). Spiroplasma citri gen. and sp. n.: A Mycoplasma-Like Organism Associated with 'Stubborn' Disease of Citrus. Int J Syst Bacteriol 23: 191–204.

Saillard C, Vignault JC, Bove JM, Raie A, Tully JG, Williamson DL, et al. (1987). Spiroplasma phoeniceum sp. nov., a New Plant-Pathogenic Species from Syria. Int J Syst Bacteriol 37: 106–115.

Salem H, Florez L, Gerardo N, Kaltenpoth M. (2015a). An out-of-body experience: the extracellular dimension for the transmission of mutualistic bacteria in insects. Proc Biol Sci 282: 20142957.

Salem H, Onchuru TO, Bauer E, Kaltenpoth M. (2015b). Symbiont transmission entails the risk of parasite infection. Biol Lett 11: 20150840.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12: 87.

Sasaki T, Kawamura M, Ishikawa H. (1996). Nitrogen recycling in the brown planthopper, Nilaparvata lugens: Involvement of yeast-like endosymbionts in uric acid metabolism. J Insect Physiol 42: 125–129.

Schmid-Hempel P. (2005). Evolutionary ecology of insect immune defenses. Annu Rev Entomol 50: 529–551.

Schneider DI, Klasson L, Lind AE, Miller WJ. (2014). More than fishing in the dark: PCR of a dispersed sequence produces simple but ultrasensitive Wolbachia detection. BMC Microbiology 14: 121.

Schwartzman JA, Ruby EG. (2016). Stress as a normal cue in the symbiotic environment. Trends Microbiol 24: 414–424.

Schwarz RS, Teixeira ÉW, Tauber JP, Birke JM, Martins MF, Fonseca I, et al. (2014). Honey bee colonies act as reservoirs for two Spiroplasmafacultative symbionts and incur complex, multiyear infection dynamics. MicrobiologyOpen 3: 341–355.

Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068–2069.

Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E. (2010). Commensal bacteria play a role in mating preference of Drosophila melanogaster. Proc Natl Acad Sci USA 107: 20051–20056.

Sheldon BC, Verhulst S. (1996). Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. Trends Ecol Evol 11: 317–321.

Shigenobu S, Stern DL. (2012). Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. Proc R Soc Lond B 280: 20121952.

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.

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.

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

Solano P, Salou E, Rayaisse J-B, Ravel S, Gimonneau G, Traore I, et al. (2015). Do tsetse flies only feed on blood? MEEGID 36: 184–189.

Starr MP, Chatterjee AK. (1972). The genus Erwinia: enterobacteria pathogenic to plants and animals. Annu Rev Microbiol 26: 389–426.

Staudacher H, Kaltenpoth M, Breeuwer JAJ, Menken SBJ, Heckel DG, Groot AT. (2016). Variability of bacterial communities in the moth Heliothis virescens indicates transient association with the host. PLoS ONE 11: e0154514.

Stevanovic AL, Arnold PA, Johnson KN. (2015). Wolbachia-mediated antiviral protection in Drosophila larvae and adults following oral infection. Appl Environ Microbiol 81: 8215–8223.

Stoll S, Feldhaar H, Fraunholz MJ, Gross R. (2010). Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant Camponotus floridanus. BMC Microbiology 10: 308.

Sudakaran S, Retz F, Kikuchi Y, Kost C, Kaltenpoth M. (2015). Evolutionary transition in symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced diet. ISME J 9: 2587–2604.

Tabata J, Hattori Y, Sakamoto H, Yukuhiro F, Fujii T, Kugimiya S, et al. (2011). Male killing and incomplete inheritance of a novel Spiroplasma in the moth Ostrinia zaguliaevi. Microb Ecol 61: 254–263.

Teixeira L, Ferreira Á, Ashburner M. (2008). The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. PLoS Biol 6: e1000002.

Telleria EL, Benoit JB, Zhao X, Savage AF, Regmi S, e Silva TLA, et al. (2014). Insights into the trypanosome-host interactions revealed through transcriptomic analysis of parasitized tsetse fly salivary glands. PLoS Negl Trop Dis 8: e2649.

Thao ML, Gullan PJ, Baumann P. (2002). Secondary (-Proteobacteria) endosymbionts infect the primary (-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl Environ Microbiol 68: 3190–3197.

Thomas DD. (2000). Insect population control using a dominant, repressible, lethal genetic system. Science 287: 2474–2476.

Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, et al. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. Genome Research 16: 149–156.

Truman JW, Riddiford LM. (1999). The origins of insect metamorphosis. Nature 401: 447–452.

Tsiropoulos GJ. (1983). Microflora associated with wild and laboratory reared adult olive fruit flies, Dacus oleae (Gmel.). J Appl Entomol 96: 337–340.

Tsiropoulos GJ. (1977). Reproduction and survival of the adult Dacus oleae feeding on pollens and honeydews. Environ Entomol 6: 390–392.

Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T. (2002). Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, Acyrthosiphon pisum. Mol Ecol 11: 2123–2135.

Tzanakakis M. (2003). Seasonal development and dormancy of insects and mites feeding on olive:a review. Neth J Zool 52: 87–224.

Tzanakakis ME, Stavrinides AS. (1973). Inhibition of development of larvae of the olive fruit fly, dacus oleae (diptera: tephritidae), in olives treated with streptomycin. Entomol Exp Appl 16: 39–47.

Vale PF, Jardine MD. (2015). Sex-specific behavioural symptoms of viral gut infection and Wolbachia in Drosophila melanogaster. J Insect Physiol 82: 28–32.

Vontas JG, Hejazi MJ, Hawkes NJ, Cosmidis N, Loukas M, Hemingway J. (2002). Resistance-associated point mutations of organophosphate insensitive acetylcholinesterase, in the olive fruit fly Bactrocera oleae. Insect Mol Biol 11: 329–336.

Vorburger C, Ganesanandamoorthy P, Kwiatkowski M. (2013). Comparing constitutive and induced costs of symbiont-conferred resistance to parasitoids in aphids. Ecol Evol 3: 706–713.

Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, et al. (2011). The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. Nature 476: 450–453.

Wamwiri FN, Alam U, Thande PC, Aksoy E, Ngure RM, Aksoy S, et al. (2013). Wolbachia, Sodalis and trypanosome co-infections in natural populations of Glossina austeni and Glossina pallidipes. Parasites & Vectors 6: 232.

Wang J, Weiss BL, Aksoy S. (2013). Tsetse fly microbiota: form and function. Front Cell Infect Microbiol 3: 69.

Wang W. (2004). A spiroplasma associated with tremor disease in the Chinese mitten crab (Eriocheir sinensis). Microbiology 150: 3035–3040.

Wang W, Gu W, Ding Z, Ren Y, Chen J, Hou Y. (2005). A novel Spiroplasma pathogen causing systemic infection in the crayfish Procambarus clarkii(Crustacea: Decapod), in China. FEMS Microbiol Lett 249: 131–137.

Wang W, Gu W, Gasparich GE, Bi K, Ou J, Meng Q, et al. (2011). Spiroplasma eriocheiris sp. nov., associated with mortality in the Chinese mitten crab, Eriocheir sinensis. Int J Syst Evol Microbiol 61: 703–708.

Wang Y, Stieglitz KA, Bubunenko M, Court DL, Stec B, Roberts MF. (2007). The structure of the R184A mutant of the inositol monophosphatase encoded by suhB and implications for its functional interactions in Escherichia coli. J Biol Chem 282: 26989–26996.

Weeks AR, Velten R, Stouthamer R. (2003). Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. *Proc Biol Sci* 270: 1857–1865.

Weiss B, Aksoy S. (2011). Microbiome influences on insect host vector competence. Trends Parasitol 27: 514–522.

Weiss BL, Maltz M, Aksoy S. (2012). Obligate symbionts activate immune system development in the tsetse fly. The Journal of Immunology 188: 3395–3403.

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

Welburn SC, Arnold K, Maudlin I, Gooday GW. (2009). Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. Parasitology 107: 141–145.

Welburn SC, Fèvre EM, Coleman PG, Odiit M, Maudlin I. (2001). Sleeping sickness: a tale of two diseases. Trends Parasitol 17: 19–24.

Welburn SC, Maudlin I. (1991). Rickettsia-like organisms, puparial temperature and susceptibility to trypanosome infection in Glossina morsitans. Parasitology 102: 201–206.

Welburn SC, Maudlin I. (1999). Tsetse–trypanosome interactions: rites of passage. Parasitology Today 15: 399–403.

Whitcomb RF, Chen TA, Williamson DL, Liao C, Tully JG, Bove JM, et al. (1986). Spiroplasma kunkelii sp. nov.: characterization of the etiological agent of Corn Stunt Disease. Int J Syst Bacteriol 36: 170–178.

Wilgenbusch JC, Swofford D. (2002). Inferring evolutionary trees with PAUP*. Curr Protoc Bioinformatics 6.4: 6.4.1–6.4.28.

Williamson DL, Sakaguchi B, Hackett KJ, Whitcomb RF, Tully JG, Carle P, et al. (1999). Spiroplasma poulsonii sp. nov., a new species associated with male-lethality in Drosophila willistoni, a neotropical species of fruit fly. Int J Syst Bacteriol 49: 611–618.

Wong ACN, Chaston JM, Douglas AE. (2013). The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis. ISME J 7: 1922–1932.

Xavier JB. (2011). Social interaction in synthetic and natural microbial communities. Mol Syst Biol 7: 483–483.

Xi Z, Khoo CCH, Dobson SL. (2005). Wolbachia establishment and invasion in an Aedes aegypti laboratory population. Science 310: 326–328.

Xie J, Butler S, Sanchez G, Mateos M. (2013). Male killing Spiroplasma protects Drosophila melanogaster against two parasitoid wasps. Heredity 112: 399–408.

Xie J, Vilchez I, Mateos M. (2010). Spiroplasma bacteria enhance survival of Drosophila hydei attacked by the parasitic wasp Leptopilina heterotoma. PLoS ONE 5: e12149.

Xin Z, Chen J. (2012). A high throughput DNA extraction method with high yield and quality. Plant Methods 8: 26.