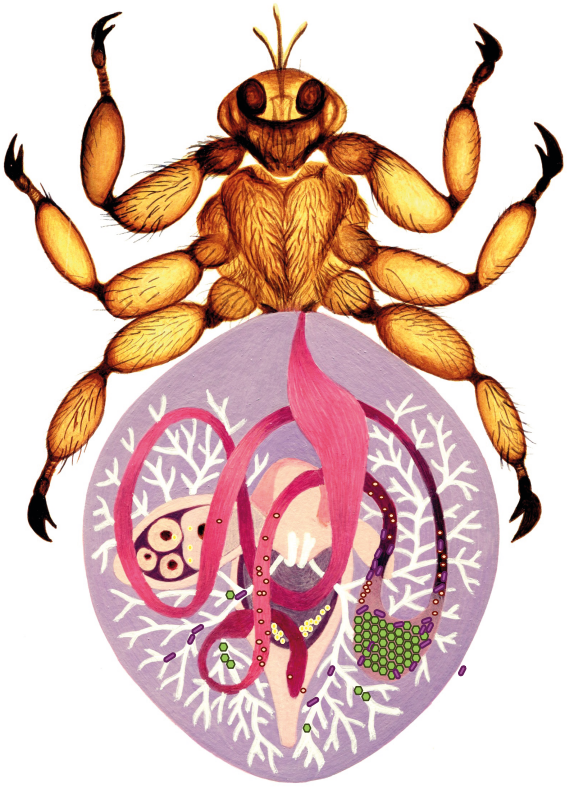
University of South Bohemia in České Budějovice

Faculty of Science



Insect symbiosis: insights into ecology, phylogenetic diversity and evolutionary dynamics

Habilitation Thesis

Eva Nováková, PhD.

2018

CONTENTS

ACKNOWLEDGEMENTS 1
ABBREVIATIONS 2

PREFACE 3

Chapter 1: The outstanding Aphids 5
List of references for Chapter 1 8

Chapter 2: Symbiosis now and then 10
Complexity of obligate mutualism: Lipoptena, Melophagus *and* Glossina *symbiosis* 10
Arsenophonus *and* Sodalis *bacteria* 12
Escaping the rabbit hole consequences: transitions, gains, losses and replacements 13
The microbiome concept 13
List of references for Chapter 2 15

LIST OF PUBLICATIONS PRESENTED IN THE THESIS 18

ATTACHED PUBLICATIONS 19

ACKNOWLEDGEMENTS

There is many people that contributed to my scientific endeavors and shaped my career to the current stage. To all of them, I owe many thanks. My special appreciations belong to my PhD advisors Prof. Vaclav Hypsa, University of South Bohemia and Prof. Nancy Moran, University of Texas at Austin. While Vaclav taught me how to do science, Nancy introduced me the world leading scientific environment. Moreover, she gave me the confidence that I can be a successful scientist and a good mother to my son. I am very grateful I had a chance to work with these outstanding people and all the members of their research teams who have become my very good friends. Many thanks also belong to the members of my developing research group for their enthusiasm for insect symbiosis and their attitude to new methodological challenges. The last but not least, I am grateful to my two men for their support, love and joy they have been giving me.

ABBREVIATIONS

AMPs	Antimicrobial peptides
Cy3	Cyanine-3
Cy5	Cyanine-5
dsRNA	double stranded RNA
FISH	Fluorescent in situ hybridization
Flc	Fluorescein
GC	Guanine+Cytosine (G+C) content of DNA
HGT	Horizontal gene transfer
HPLC	High-performance liquid chromatography
IMG	Integrated Microbial Genomes & Microbiomes
Kb	Kilobase
Mb	Megabase
NMDS	Non-Metric Multidimensional Scaling
P-symbiont	Primary symbiont
RNAi	RNA interference
S-symbiont	Secondary symbiont
siRNA	Short interfering RNA
WNV	West Nile Virus

PREFACE

Host-parasite associations keep the central place in evolutionary ecology. However, to uncover the details of the genetic architecture underlying these associations, one cannot limit the study to the two obvious counterparts. The evolution of many parasitic groups, especially blood sucking arthropods, have been shaped by symbiosis. Alliances with bacteria lead to metabolic integration that enabled hematophagous parasites to flourish on nutritionally restricted vertebrate blood. While in some groups, bacterial symbionts promoted evolution of their host towards parasitism, in others they contribute to their host resistance against parasites. In addition, bacterial symbionts or whole symbiotic communities of insect vectors modulate their vectorial capacity for transmission of different parasites and pathogens. Parasitism is simply intertangled with symbiosis on different levels and, by a more general view of a common interaction between unrelated organisms, even understood as a form of symbiosis itself.

Since my first research steps in Vaclav Hypsa's laboratory, my work has thus focused on genetic architecture and evolution of bacterial symbiosis. While primarily oriented on blood sucking hosts, the career line took a slight, but extremely useful, detour during my stays with Nancy Moran, the leading scientist in the field, where I discovered the magic world of sap sucking hosts and their symbionts. Similar to the blood sucking insects, aphids are a priceless model in which one can address questions on horizontal gene transfer as a crucial mechanism of evolution, or investigate the coevolutionary impact of symbiont gain, spread and loss. My close encounters with this sap feeding model are referred to in the chapter entitled **"The Outstanding Aphids"**. The word *outstanding* here points out not only to the intriguing evolution and ecology of these insects, and thus an endless opportunity for a new research enterprise, but also it highlights my own perception of the aphid research throughout my career. Aphids took me not only far behind the symbiotic association itself, but lately also offered a methodological outreach into experimental approaches based on RNA mediated manipulation of gene expression.

The winding path towards my own research group, broadly oriented on an interplay among blood sucking vectors, their symbionts and transmitted parasites/pathogens, is then overviewed in the second chapter called **"Symbiosis now and then".** The title refers to the methodological development that has enabled our understanding of complex symbiotic systems. To some degree, it also refers to a career flashback I have encountered. The very first hosts I explored in my undergraduate years under Vaclav's supervision were the blood sucking kissing bugs (Reduviidae: Triatominae). A year ago, I came back to the very same system, for which metagenomics and transcriptomics opened entirely new entryways, and since Triatominae microbiomes has become the principal topic for the group (see the section ***"The microbiome concept"***).

The papers selected for this thesis are not ordered chronologically according to the year of publication. They rather accompany and broaden the content of the two chapters introducing complex biological characteristics of insect hosts, the allied bacteria and parasites/pathogens.

My further contributions on arthropod vectored parasites/pathogens constitute the two recent manuscripts[1, 2] that are however not further elaborated in this thesis.

[1] Pistone D, Pajoro M, Novakova E, Vicari N, Gaiardelli C, Viganò R, et al. 2017. Ticks and bacterial tick-borne pathogens in Piemonte region, Northwest Italy. Experimental and Applied Acarology, *73*: 477-491.

[2] Păstrav IR, Ionică AM, Peştean C, Novakova E., Modrý D, Mihalca AD, 2018. Peripheral venous vs. capillary microfilariaemia in a dog co-infected with *Dirofilaria repens* and *D. immitis*: a comparative approach using triatomine bugs for blood collection. Veterinary Parasitology, *257:* 54-57.

Chapter 1: The Outstanding Aphids

Aphids (Hemiptera: Aphididae) are one of the most diverse insect taxa in temperate regions feeding on plant phloem. More than 4400 species (Blackman & Eastop, 1994) feature an immense variability of biological traits including color polymorphism, complicated lifecycles with sexual and asexual reproduction, various female forms and host-plant alternation. All these are underlined by a complex and, to some extent, very unique evolutionary history. The crucial event, predating the origin of this group, corresponds to an ancient bacterial infection resulting in the establishment of very intimate symbiotic alliance. The acquisition of *Buchnera aphidicola*, dated to 150 MYA (Munson et al., 1991; Moran et al., 1993), was followed by a strict co-speciation between the host and the symbiont (Moran et al., 1993). The common evolutionary history, confirmed by mirroring phylogenies available for some aphid genera and their *Buchnera* symbionts (e.g. Clark et al., 2000; Liu et al., 2013), offers an extra value to the studies focused on host evolution. It allows for treating the symbiotic genome in the same way as mitochondria, and thus provides an alternative genetic information to the host nuclear genes.

While suggested to be applied for several other insect hosts (e.g. Kölsch and Pedersen, 2010), this approach was particularly useful for untangling evolutionary history of aphids that underwent a rapid adaptive radiation (von Dohlen and Moran, 2000). As a result of such a rapid adaptive event, aphid mitochondria retained a low level of phylogenetic signal and the nuclear genes experienced number of duplications resulting in extremely high number of paralogs hampering the molecular analyses (Rispe et al., 2007; Novakova and Moran 2012). Thus, based on aphid markers, the relationships among many subfamilies and tribes had for long remained poorly resolved and the absence of a reliable phylogeny obstructed meaningful interpretation of their biological traits in the evolutionary context. Taking advantage of a high number of *Buchnera aphidicola* cells in an aphid individual and their extremely streamlined bacterial genomes (416-644 kb) that naturally lack introns and gene duplications, we have elucidated some of notoriously problematic phylogenetic relationships within Aphididae (Novakova et al., 2013, *PAPER 1*). We have confirmed monophyly of individual tribes, monophyly of Lachninae and monophyly of Aphidinae (including Aphidini, Macrosiphini, Pterocommatini). Paraphyletic nature of Macrosiphini was fully revealed with respect to Pterocommatini. Although the five bacterial markers we used for an extended taxonomic set of 70 aphid species helped corroboration and/or rejection of particular phylogenetic hypotheses, the complete resolution of aphid phylogeny could have not been reached (Novakova et al., 2013).

The establishment of a symbiotic relationship with bacteria, introduced above, is almost ubiquitous among insects. Aphids, however, experienced another outstanding evolutionary event with major functional consequences. A horizontal gene transfer (HGT) of carotenogenic genes from a fugal donor predated the split of aphids (Aphididae) and adelgids (Adelgidae) (Novakova and Moran, 2012, *PAPER 2*). Many genomes of Bacteria, Archaea, unicellular eukaryotes and plants code for biosynthesis of carotenoids that confer variety of benefits to these organisms. However, the presence of the core enzymatic machinery for carotenoid biosynthesis in the genomes of aphids makes them exceptional among animals (Moran and Jarvik, 2010; Novakova and Moran, 2012). Although the presence of carotenoids in bright colored aphid bodies was apparent, for many decades, the capacity of the aphid genome for carotenoid production had remained unrecognized. The carotenoid origin had been initially

assigned to the plant sap diet or alternatively to the aphid bacterial symbionts (e.g. Jenkins et al., 1999). While the first hypothesis had flaws from its very beginning, when the carotenoid profiles of aphids were shown to differ dramatically from those in their host plants (Czeczuga, 1976), the second, symbiont-based explanation, could have been rejected much later with the first genome sequences of *Buchnera* and other facultative associates of aphids (Shigenobu et al., 2000; Degnan et al., 2009). Only then, and very unexpectedly, multiple enzymes for carotenoid biosynthesis have been found to be encoded by the pea aphid (*Acyrthosiphon pisum*) genome released in 2010 (International Aphid Genomics Consortium, 2010; Moran and Jarvik, 2010). Shortly after this breakthrough, we have provided a complex picture of the carotenogenic genes evolution in aphids (Novakova and Moran, 2012, *PAPER 2*). We have narrowed down the origin of these horizontally acquired genes to a fungal donor closely related to Mucoromycotina. Furthermore, we have shown that the carotene desaturase gene underwent multiple duplication and recombination events followed by bouts of positive selection that resulted in a wide variety of carotenoid profiles in different aphid species (Figure 1).

Figure 1. Examples of HPLC based carotenoid profiles obtained for different aphid species from three tribes. The bar width indicates proportional representation among carotenoids detected in selected species. The drawings on the right represent approximate color in life. Xanthins of likely plant origin are in green colors (altered from Novakova and Moran, 2012)

The research initiative targeting aphids, particularly the model species *Acyrthosiphon pisum*, stems not only from the fascinating evolutionary history of this group but also from the impact the insects pose on the world agriculture production. Since my general interest lies in understanding fundamental evolutionary and molecular mechanisms shaping various symbiotic systems, my contribution to mechanisms of RNA interference in aphids was not primarily meant

to gain applicable results for pest control strategies. The major motivation was rather a common need for new tools in functional genomics that could, at the first place, validate computationally predicted gene annotations in numerous insect genomes and, second, enable reliable and reproducible gene manipulation.

RNA interference (RNAi) is a mechanism of posttranscriptional gene silencing. Promoted either by double stranded or short interfering RNA molecules (dsRNA or siRNA), the RNAi pathway targets particular transcripts (mRNA) and proceed with their degradation. In aphids, the success of RNAi for gene silencing have been highly variable, from a high efficacy reported in green peach aphids (Guo et al., 2014; Coleman et al., 2015) to inconsistent effects in pea aphids (e.g. Mao and Zeng, 2012; Christiaens et al., 2014). Because the delivery of RNA molecules into the experimental insects has been identified as one of the main obstacles, we have focused on a novel technique using aerosolized siRNA-nanoparticle complexes (Thairu et al., 2017, *PAPER 3*). While, in general, the siRNA-nanoparticle complexes significantly increased the efficacy of gene knockdown (compered to sole siRNA molecules), we have still encountered variable success rate among the three tested aphid species. While modest for *Acyrthosiphon pisum* and *Shizaphis graminum*, in *Aphis glycines, blt* gene knockdown lead to a phenotypic change (Thairu et al., 2017).

List of references for Chapter 1

Blackman, R. L., & Eastop, V. F. (1994). Aphids on the world's trees: an identification and information guide. *CAB International, Wallingford, United Kingdom*.

Christiaens, O., Swevers, L., & Smagghe, G. (2014). DsRNA degradation in the pea aphid (Acyrthosiphon pisum) associated with lack of response in RNAi feeding and injection assay. *Peptides*, *53*, 307-314.

Clark, M. A., Moran, N. A., Baumann, P., & Wernegreen, J. J. (2000). Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution*, *54*(2), 517-525.

Coleman, A. D., Wouters, R. H. M., Mugford, S. T., & Hogenhout, S. A. (2014). Persistence and transgenerational effect of plant-mediated RNAi in aphids. *Journal of experimental botany*, *66*(2), 541-548.

Czeczuga, B. (1976). Investigations on the carotenoids in nineteen species of aphids and their host plants. *ZP Zool Pol*.

Degnan, P. H., Yu, Y., Sisneros, N., Wing, R. A., & Moran, N. A. (2009). *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, *106*(22), 9063-9068.

Jenkins, R. L., Loxdale, H. D., Brookes, C. P., & Dixon, A. F. G. (1999). The major carotenoid pigments of the grain aphid, *Sitobion avenae* (Hemiptera: Aphididae). *Physiological Entomology*, *24*(2), 171-178.

Guo, H., Song, X., Wang, G., Yang, K., Wang, Y., Niu, L., ... & Fang, R. (2014). Plant-generated artificial small RNAs mediated aphid resistance. *PLoS One*, *9*(5), e97410.

International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS biology*, *8*(2), e1000313.

Kölsch, G., & Pedersen, B. V. (2010). Can the tight co-speciation between reed beetles (Col., Chrysomelidae, Donaciinae) and their bacterial endosymbionts, which provide cocoon material, clarify the deeper phylogeny of the hosts?. *Molecular phylogenetics and evolution*, *54*(3), 810-821.

Liu, L., Huang, X., Zhang, R., Jiang, L., & Qiao, G. (2013). Phylogenetic congruence between *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect–bacteria parallel evolution. *Systematic Entomology*, *38*(1), 81-92.

Mao, J., & Zeng, F. (2012). Feeding-based RNA interference of a gap gene is lethal to the pea aphid, *Acyrthosiphon pisum*. *PLoS One*, *7*(11), e48718.

Moran, N. A., Munson, M. A., Baumann, P., & Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B*, *253*(1337), 167-171.

Moran, N. A., & Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, *328*(5978), 624-627.

Munson, M. A., Baumann, P., Clark, M. A., Baumann, L., Moran, N. A., Voegtlin, D. J., & Campbell, B. C. (1991). Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *Journal of bacteriology*, *173*(20), 6321-6324.

Nováková, E., & Moran, N. A. (2012). Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Molecular biology and evolution*, *29*(1), 313-323.

Nováková, E., Hypša, V., Klein, J., Foottit, R. G., von Dohlen, C. D., & Moran, N. A. (2013). Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution*, *68*(1), 42-54.

Rispe, C., Kutsukake, M., Doublet, V., Hudaverdian, S., Legeai, F., Simon, J. C., ... & Fukatsu, T. (2007). Large gene family expansion and variable selective pressures for cathepsin B in aphids. *Molecular biology and evolution*, *25*(1), 5-17.

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*(6800), 81.

Thairu, M. W., Skidmore, I. H., Bansal, R., **Nováková, E.**, Hansen, T. E., Li-Byarlay, H., ... & Hansen, A. K. (2017). Efficacy of RNA interference knockdown using aerosolized short interfering RNAs bound to nanoparticles in three diverse aphid species. *Insect molecular biology*, *26*(3), 356-368.

Von Dohlen C. D., & Moran, N. A. (2000). Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biological Journal of the Linnean Society*, *71*(4), 689-717.

Chapter 2: Symbiosis now and then

In the field of insect symbiosis, "the outstanding" aphids and *Buchnera aphidicola* have become a model system for which the evolutionary and functional interdependencies have been thoroughly studied since the dawn of molecular biology. The system represents an iconic example of symbiosis between a host and a coevolving obligate mutualist, or so called primary (P-) symbiont (e.g. reviewed in Shigenobu et al., 2011; Douglas, 2016). The terms P-symbiont and S-symbiont (primary and secondary), coined by the pioneer of insect symbiosis Paul Buchner (1965), founded a firm but very simplified frame. P-symbiont, as the evolutionary original symbiont, transmits vertically through the host generations, provides essential nutrients to the host, and resides in its specialized cells. The term S-symbiont encompasses bacteria in various host tissues with diverse relationships to the hosts in which populations S-symbionts utilize both vertical and horizontal transfer. Since the terminology appears in many of the discussed manuscripts and provides a transparent concept to a broader reader audience, I have decided to note the above short explanations.

However, already by the end of the last century, the P- and S-symbiont categories were proven as extremely insufficient, failing to encompass the growing diversity and dynamics of known symbiotic associations. This can be well illustrated on symbiosis established by number of hematophagous insects that rely on sufficient levels of B vitamins provided by the allied bacteria. The relationships display considerable degree of variability spanning from the least intimate with gut microbial communities, e.g. in kissing bugs (Rodriguez-Ruano et al., 2018), to highly specialized interactions with various complexity, found in lice, bed bugs, tsetse flies, louse flies and bat flies (Aksoy et al., 1997; Sasaki-Fukatsu et al., 2006; Hypsa and Krizek, 2007; Hosokawa et al., 2010; Novakova et al., 2015). Particularly Hippoboscoidea, including tsetse flies (Glossinidae), louse flies and bat flies (Hippoboscidae, Nycteribidae and Streblidae), offers a great insight into the complexity of arrangements, genomic modifications, and metabolic interconnections found in symbiotic systems.

Complexity of obligate mutualism: Lipoptena, Melophagus *and* Glossina *symbiosis*

The deer ked, *Lipoptena cervi,* and the sheep ked, *Melophagus ovinus* (Hippoboscidae: Lipopteninae), are two closely related species of parasitic flies with very similar biology. Yet, the symbioses they maintain with bacteria substantially differ. While the bacteriome (the specialized host symbiotic organ) of *Lipoptena cervi* houses a single bacterial mutualist, *Candidatus* Arsenophonus lipopteni (Figure 2A; Novakova et al., 2016, *PAPER 4*), the symbiosis maintained by *Melophagus ovinus* is much more complex, involving four unrelated bacteria (Novakova et al., 2015, *PAPER 5*). In this case, the bacteriome is inhabited by a typical obligate mutualist, *Candidatus* Arsenophonus melophagi, residing inside the symbiotic cells (bacteriocytes) and a S-symbiont, *Candidatus* Sodalis melophagi, restricted into the inter-cellular space (Figure 2B). In the posterior midgut, however, *Ca.* Sodalis melophagi enters the epithelial cells which then resemble the bacteriocytes (Figure 2B and 2C). Furthermore, the system encompasses bacteria of the genus *Wolbachia* located in various tissues, including the gut epithelia, and a putative sheep parasite, *Bartonella melophagi,* in the gut lumen (Figure 2C; Novakova et al., 2015).

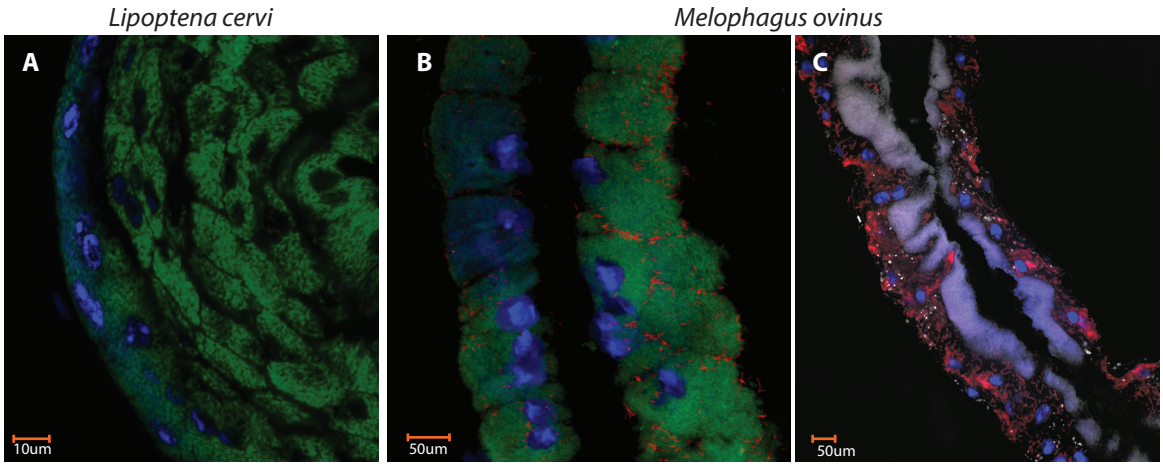


Figure 2. Visualization of different symbiotic arrangement in *Lipoptena cervi* and *Melophagus ovinus* by 16S rRNA-targeted fluorescent in situ hybridization (FISH) using multiple labeled probes: Flc, Cy3, Cy5. *Lipoptena cervi* bacteriome harboring *Ca.* Arsenophonus lipopteni targeted with green (Flc) labeled probe **(A)**. *Melophagus ovinus* bacteriome with *Ca.* Arsenophonus melophagi in green (Flc) within the bacteriocytes and *Ca.* Sodalis melophagi in red (Cy3) restricted to the intercellular space **(B)**. *Ca.* Sodalis melophagi (red; Cy3) and *Wolbachi*a sp. (white; Cy5) in the midgut section. The nonspecific signal in the gut lumen was produced by *Bartonella melophagi* cells **(**Texas Red targeted FISH not shown; **C).** The blue signal corresponds to DAPI stained host nuclei (adopted from Novakova et al., 2015 and 2016)

More interestingly, within the frame of higher taxonomical level (i.e. Hippoboscoidea), the sheep ked (Hippoboscidae) symbiosis provides a remarkable analogy to the association between tse-tse fly (Glossinidae) and its symbiotic bacteria (Novakova et al., 2015). Both hosts established symbiosis with an obligate mutualist and two facultative intracellular associates, *Sodalis* and *Wolbachia*. The two obligate mutualists however originate in phylogenetically distant bacterial genera (*Wigglesworthia* and *Arsenophonus*), and their genomes differ considerably in size (0.72 Mb and 1.16 Mb, respectively). While the entire symbiotic systems feature apparent functional convergencies, they also display remarkable differences found among their counterparts. For instance, illustrated on B vitamins biosynthesis which is traditionally considered a core aspect of symbioses in blood-sucking insect (e.g. Nogge, 1981; Hosokawa et al, 2010; Nikoh et al., 2014), thiamine production and transport rely on different players. In the sheep ked, thiamine biosynthesis relies on *Sodalis melophagi* and/or *Bartonella melophagi*, whereas the obligate *Arsenophonus* symbionts can only transport this compound (Novakova et al., 2015). In contrary, the tsetse fly obligate mutualist, *Wigglesworthia glossinidia*, codes for thiamine biosynthesis and *Sodalis glossinidius* only possesses thiamine specific transporters (Akman et al., 2002; Toh et al., 2006; Belda et al., 2010).

Arsenophonus *and* Sodalis *bacteria*

The genus *Arsenophonus* and *Sodalis* bacteria exhibit indeed a stunning diversity of relationships they engaged in with a broad spectrum of organisms, including vertebrates and plants (e.g. Novakova and Hypsa, 2007, *PAPER 6*; Hypsa and Novakova, 2009, *BOOK CHAPTER 1*; Novakova et al., 2009, *PAPER 7*; Wilkes et al., 2011, *BOOK CHAPTER 2*; Chrudimsky et al., 2012, *PAPER 8*; Novakova et al., 2015; Novakova et al., 2016). Along with *Wolbachia* symbionts (Alphaproteobacteria), undoubtedly representing the most widespread bacterial associates of arthropods (e.g. Werren et al., 2008), the gammaproteobacterial genera *Arsenophonus* and *Sodalis* constitute a great proportion of symbiotic diversity (Wilkes et al., 2011; Santos-Garcia et al., 2017). While some *Arsenophonus* strains express a harmful phenotype as, for instance, the son-killers of parasitic wasps (Gherna et al., 1991) or phytopathogens of crop, others, found in various tissues of many hemipteran and dipteran species, do not pose any apparent effect on the host fitness (e.g. Novakova et al., 2009; Wilkes et al., 2011). In several blood sucking insects, as described above, *Arsenophonus* bacteria play the role of indispensable obligate symbionts providing essential nutrients (e.g. Allen et al., 2007; Novakova et al., 2015; Novakova et al., 2016; Boyd et al., 2017). Biology of the genus *Arsenophonus* is reviewed in the second book chapter accompanying this thesis.

A comparable range of different ecological niches and specific effects on the hosts has been found for *Sodalis* bacteria. Since the description of the type species, *Ca.* Sodalis glossinidius, from tse-tse flies, number of various *Sodalis* symbionts have been recorded from weevils, hippoboscid louse fly, chewing lice and seal lice, aphids, psyllids, scale insects, spittlebugs, and stinkbugs (reviewed in Santos-Garcia et al., 2017). A member of the genus *Sodalis* that possibly represents the free living seed of the above mentioned symbioses was isolated from a human wound infection (Clayton et al., 2012).
Such distinctions in the phenotypic characteristics are naturally well reflected in the genome contents of these bacteria which size and the GC content span the range from several thousand to little over a hundred kilobases, and from over 50% of GC to as low as 17.1% in the smallest known genome of *Nasuia deltocephalinicola* (Bennett and Moran, 2013). For instance, these major genomic characteristics for known *Arsenophonus* genomes fall between 4029 Kb of 38% GC and 574 Kb of 28% GC, for the son-killing *Arsenophonus nasoniae* and the nutritional mutualist *Riesia pediculicola*, respectively (IMG genome IDs: 2512564001 and 2510065001). Similarly, genomes of *Sodalis* symbionts feature a very close range (e.g. Santos-Garcia et al., 2017).

While the genomes of S-symbionts evidence ongoing dynamics, including both gene gains and gene losses, the genomes of obligate mutualists have been shaped, to a different extent, by degradation processes. Promoted by a strong genetic drift acting in small symbiont populations that undergo repeated bottlenecks on one hand, and lack efficient selection and recombination on the other, their genomes face fixation of deleterious mutations, rapid protein evolution and an overall extreme reduction. Consequently, the host, in a vital need for maintaining the symbiosis, evolves adaptive compensations, including gain of bacterial genes that aid control of the symbiotic relationship (Husnik et al., 2013; Nakabachi et al., 2014; Sloan et al., 2014). The symbiotic systems thus face dependences that evolved beyond the original benefits and pose potential negative risks. For instance, symbiosis driven immune-system modifications may result in fitness costs (e.g. Gerardo et al., 2010; Douglas et al., 2011) and/or radical degeneration of symbiont genomes may extremely limit the ecological range of host species

and thus increase the risk of extinction (Bennett and Moran, 2015). The commitment of the symbiotic counterparts to such perilous obligate alliance was described by Nancy Moran as "the evolutionary rabbit hole of obligate symbiosis, implying a generally irreversible journey into a very odd world where the usual rules do not apply" (Bennett and Moran, 2015).

Escaping the rabbit hole consequences: transitions, gains, losses and replacements

The known mosaic of symbiotic associations in numerous host groups thus corresponds to a present-day window into the evolution throughout which the bacteria are being gained, lost or replaced in the symbiotic systems that seek their way out of the spiral leading to the symbiotic death. This evolutionary ultimatum is most commonly broken by adding a new bacterial associate which complements the current arrangement of the symbiotic system or eventually replace the original P-symbiont. Such a scenario could be exemplified on symbiosis of bloodsucking louse flies (Hippoboscidae) affected by frequent replacements of their bacterial partners (Sochova et al., 2017, *PAPER 9*). The *Arsenophonus* and *Sodalis* infections seem to be the most common seeds for obligate symbioses that have evolved independently several times in this host group. According to our recent results, there are at least four lineages of likely obligate *Arsenophonus* endosymbionts. *Sodalis*-like obligate endosymbionts however form a monophyletic clade with likely a single origin in one of the Hippoboscidae tribes (Sochova et al., 2017).

For some bacterial associates, transitions towards obligate symbiosis is a gradual process of genome reduction, whereas others encounter a crucial evolutionary event of gene acquisition promoting their symbiotic role. Such a case applies for a lineage of *Legionella* that acquired horizontally transferred biotin operon and gave rise to the only known symbiotic lineage of Legionellaceae (Rihova et al., 2017, *PAPER 10*). *Candidatus* Legionella polyplacis, the obligate symbiont of *Polyplax* lice, then underwent a surprisingly convergent evolution leading to the genome characteristics that closely resemble those of unrelated enterobacterial symbionts from the hominid lice. Compared to the 38 known *Legionella* genomes which sizes vary between 2 and 5 Mb (Burstein et al., 2016), *Ca.* Legionella polyplacis experienced considerable reduction, bearing only 484 coding genes in 0.53 Mb long genome of 23%GC (Rihova et al., 2017).

The microbiome concept

The microbiome concept treats the symbiotic bacterial community as a collective genome of microorganisms which share a common environment and are supposedly interconnected by metabolic and/or genetic interactions (Lederberg and McCray, 2001). However, this view could have only been reached with the boom of high throughput sequencing techniques and fully flourished during the last decade. The technical development combined with the new theoretical concept enable us to answer various questions on interactions of microbiomes with the host genetic background, significance of the host humoral and cellular factors in shaping the community, the roles of individual defense molecules (AMPs, lysozyme or prophenoloxidase), or impact of the microbial communities on vector capacities in blood feeding disease vectors.

The two last studies of this thesis embarked on the microbiome concept exploring the interplay of host-symbiotic bacteria-and transmitted parasites/pathogens in two groups of blood sucking insect, mosquitoes and kissing bugs (Novakova et al., 2017, *PAPER 11* and Rodriguez-Ruano et al., 2018, *PAPER 12*). The microbial communities were characterized for adult individuals from 11 mosquito species sampled over 3 years in six regions of southern Ontario, Canada. The host species was identified as the most important determinant shaping the microbiome profiles. In *Culex pipiens/restuans*, the microbiome undergoes consistent seasonal compositional shifts associated with West Nile Virus (WNV) prevalence patterns. The higher temperatures seem to trigger reduction in the *Wolbachia* abundance which in turn leads to a greater susceptibility to WNV in the subsequent generation of *C. pipiens/restuans* hosts. Overall, the results suggest that microbiome dynamics may explain some of the variation in vector competence previously attributed to local environmental processes (Novakova et al., 2017).

Compared to the mosquitoes for which the microbiome based surveys have constituted a solid research body, the study on microbiomes of kissing bugs (Reduviidae: Triatominae) are still in its infancy. With the last paper, focused on Triatominae-microbiome-*Trypanosoma cruzi* interface, we provide a general guideline on how to analyze, treat, and potentially use the microbiomes taking in account their ontogenetic variability (Rodriguez-Ruano et al., 2018).

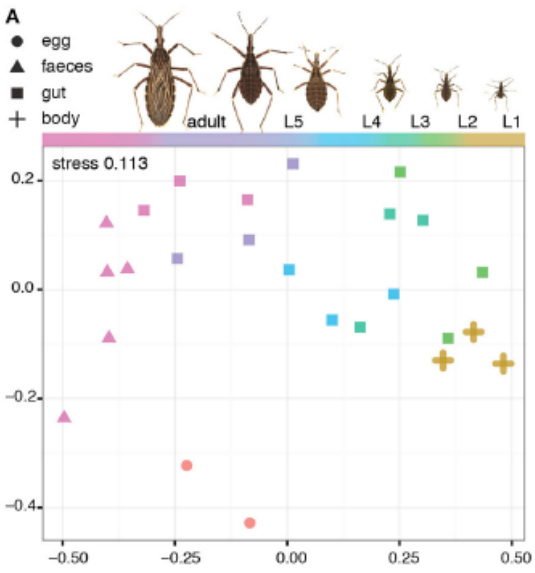


Figure 3. Visualization of the ontogenetic shift of *Rhodnius prolixus* microbiome. The plot is based on calculations of Bray-Curtis dissimilarities used in Non-Metric Multidimensional Scaling (NMDS). In two dimensional space, the closer the symbols are, the higher the overall similarity among microbiomes is (adopted from Rodriguez-Ruano et al., 2018).

In addition, this pilot study brings the very first description of microbiome profiles in wild populations of *T. sanguisuga* and *T. protracta*, the two abundant North American species able to transmit *T. cruzi*, the causative agent of Chagas disease. Interestingly, *Trypanosoma cruzi* positive *T. protracta* specimens tend to harbor a more diverse microbiome than the negative individuals, pointing out the significance of Triatominae microbiomes in the epidemiology of Chagas disease (Rodriguez-Ruano et al., 2018).

List of references for Chapter 2

Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., & Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia. Nature genetics*, *32*(3), 402.

Aksoy, S., Chen, X. A., & Hypsa, V. (1997). Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). *Insect molecular biology*, *6*(2), 183-190.

Allen, J. M., Reed, D. L., Perotti, M. A., & Braig, H. R. (2007). Evolutionary relationships of "*Candidatus* Riesia spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Applied and Environmental Microbiology*, *73*(5), 1659-1664.

Belda, E., Moya, A., Bentley, S., & Silva, F. J. (2010). Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. *BMC genomics*, *11*(1), 449.

Bennett, G. M., & Moran, N. A. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome biology and evolution*, *5*(9), 1675-1688.

Bennett, G. M., & Moran, N. A. (2015). Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences*, *112*(33), 10169-10176.

Boyd, B. M., Allen, J. M., Nguyen, N. P., Vachaspati, P., Quicksall, Z. S., Warnow, T., ... & Reed, D. L. (2017). Primates, lice and bacteria: speciation and genome evolution in the symbionts of hominid lice. *Molecular biology and evolution*, *34*(7), 1743-1757.

Buchner, P. (1965). Symbiosis in animals which suck plant juices. *Endosymbiosis of animals with plant microorganisms*, 210-432.

Burstein, D., Amaro, F., Zusman, T., Lifshitz, Z., Cohen, O., Gilbert, J. A., ... & Segal, G. (2016). Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. *Nature genetics*, *48*(2), 167.

Chrudimský, T., Husník, F., **Nováková, E.**, & Hypša, V. (2012). *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. *PLoS One*, *7*(7), e40354.

Clayton, A. L., Oakeson, K. F., Gutin, M., Pontes, A., Dunn, D. M., von Niederhausern, A. C., ... & Dale, C. (2012). A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. *PLoS genetics*, *8*(11), e1002990.

Douglas, A. E., Bouvaine, S., & Russell, R. R. (2011). How the insect immune system interacts with an obligate symbiotic bacterium. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1704), 333-338.

Douglas, A. E. (2016). How multi-partner endosymbioses function. *Nature Reviews Microbiology*, *14*(12), 731.

Gerardo, N. M., Altincicek, B., Anselme, C., Atamian, H., Barribeau, S. M., De Vos, M., ... & Heddi, A. (2010). Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome biology*, *11*(2), R21.

Gherna, R. L., Werren, J. H., Weisburg, W., Cote, R., Woese, C. R., Mandelco, L., & Brenner, D. J. (1991). *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *International Journal of Systematic and Evolutionary Microbiology*, *41*(4), 563-565.

Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X. Y., & Fukatsu, T. (2010). *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proceedings of the National Academy of Sciences*, *107*(2), 769-774.

Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., ... & Von Dohlen, C. D. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, *153*(7), 1567-1578.

Hypša, V., & Křížek, J. (2007). Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). *Microbial ecology*, *54*(2), 242-251.

Hypša, V., & **Nováková, E.** (2009). Insect symbionts and molecular phylogenetics. *Insect symbiosis*, *3*, 1-31.

Lederberg, J., & McCray, A. T. (2001). Ome SweetOmics--A Genealogical Treasury of Words. *The Scientist*, *15*(7), 8-8.

Nakabachi, A., Ishida, K., Hongoh, Y., Ohkuma, M., & Miyagishima, S. Y. (2014). Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Current Biology*, *24*(14), R640-R641.

Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., & Fukatsu, T. (2014). Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences*, *111*(28), 10257-10262.

Nogge, G. (1981). Significance of symbionts for the maintenance of an optimalnutritional state of successful reproduction in hematophagous arthropods. *Parasitology*, *82*, 299-304.

Nováková, E., & Hypša, V. (2007). A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS microbiology letters*, *269*(1), 131-135.

Nováková, E., Hypša, V., & Moran, N. A. (2009). *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC microbiology*, *9*(1), 143.

Nováková, E., Husník, F., Šochová, E., & Hypša, V. (2015). *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and environmental microbiology*, *81*(18), 6189-6199.

Nováková, E., Hypša, V., Nguyen, P., Husník, F., & Darby, A. C. (2016). Genome sequence of *Candidatus* Arsenophonus lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *Standards in genomic sciences*, *11*(1), 72.

Novakova, E., Woodhams, D. C., Rodríguez-Ruano, S. M., Brucker, R. M., Leff, J. W., Maharaj, A., ... & Scott, J. (2017). Mosquito microbiome dynamics, a background for prevalence and seasonality of West Nile virus. *Frontiers in microbiology*, *8*, 526.

Rodriguez-Ruano, S. M., Škochová, V., Rego, R. O. M., Schmidt, J. O., Roachell, W., Hypša, V., & **Novakova, E.** (2018). Microbiomes of North American Triatominae: the grounds for Chagas disease epidemiology. *Frontiers in Microbiology*, *9*, 1167.

Říhová, J., **Nováková, E**., Husník, F., & Hypša, V. (2017). *Legionella* becoming a mutualist: adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome biology and evolution*, *9*(11), 2946-2957.

Santos-Garcia, D., Silva, F. J., Morin, S., Dettner, K., & Kuechler, S. M. (2017). The All-Rounder *Sodalis*: A New Bacteriome-Associated Endosymbiont of the Lygaeoid Bug *Henestaris halophilus* (Heteroptera: Henestarinae) and a Critical Examination of Its Evolution. *Genome biology and evolution*, *9*(10), 2893-2910.

Sasaki-Fukatsu, K., Koga, R., Nikoh, N., Yoshizawa, K., Kasai, S., Mihara, M., ... & Fukatsu, T. (2006). Symbiotic bacteria associated with stomach discs of human lice. *Applied and environmental microbiology*, *72*(11), 7349-7352.

Shigenobu, S., & Wilson, A. C. (2011). Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cellular and molecular life sciences*, *68*(8), 1297-1309.

Sloan, D. B., Nakabachi, A., Richards, S., Qu, J., Murali, S. C., Gibbs, R. A., & Moran, N. A. (2014). Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular biology and evolution*, *31*(4), 857-871.

Šochová, E., Husník, F., **Nováková, E.**, Halajian, A., & Hypša, V. (2017). *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ*, *5*, e4099.

Toh, H., Weiss, B. L., Perkin, S. A., Yamashita, A., Oshima, K., Hattori, M., & Aksoy, S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome research*, *16*(2), 149-156.

Wilkes, T. E., Duron, O., Darby, A. C., Hypša, V., **Nováková, E.,** & Hurst, G. D. (2011). 12 The Genus *Arsenophonus*. *Manipulative tenants: bacteria associated with arthropods*, 225.

Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, *6*(10), 741.

LIST OF PUBLICATIONS PRESENTED IN THE THESIS

1. **Nováková, E**., Hypša, V., Klein, J., Foottit, R. G., von Dohlen, C. D., & Moran, N. A. (2013). Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution*, *68*(1), 42-54.
2. **Nováková, E.,** & Moran, N. A. (2012). Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Molecular biology and evolution*, *29*(1), 313-323.
3. Thairu, M. W., Skidmore, I. H., Bansal, R., **Nováková, E**., Hansen, T. E., Li-Byarlay, H., ... & Hansen, A. K. (2017). Efficacy of RNA interference knockdown using aerosolized short interfering RNAs bound to nanoparticles in three diverse aphid species. *Insect molecular biology*, *26*(3), 356-368.
4. **Nováková, E**., Hypša, V., Nguyen, P., Husník, F., & Darby, A. C. (2016). Genome sequence of *Candidatus* Arsenophonus lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *Standards in genomic sciences*, *11*(1), 72.
5. **Nováková, E**., Husník, F., Šochová, E., & Hypša, V. (2015). *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and environmental microbiology*, *81*(18), 6189-6199.
6. **Nováková, E.**, & Hypša, V. (2007). A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS microbiology letters*, *269*(1), 131-135.
7. **Nováková, E.**, Hypša, V., & Moran, N. A. (2009). *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC microbiology*, *9*(1), 143.
8. Chrudimský, T., Husník, F., **Nováková, E.**, & Hypša, V. (2012). *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. *PLoS One*, *7*(7), e40354.
9. Šochová, E., Husník, F., **Nováková, E.**, Halajian, A., & Hypša, V. (2017). *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ*, *5*, e4099.
10. Říhová, J., **Nováková, E**., Husník, F., & Hypša, V. (2017). *Legionella* becoming a mutualist: adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata. Genome biology and evolution*, *9*(11), 2946-2957.
11. **Novakova, E.**, Woodhams, D. C., Rodríguez-Ruano, S. M., Brucker, R. M., Leff, J. W., Maharaj, A., ... & Scott, J. (2017). Mosquito microbiome dynamics, a background for prevalence and seasonality of West Nile virus. *Frontiers in microbiology*, *8*, 526.
12. Rodriguez-Ruano, S. M., Škochová, V., Rego, R. O. M., Schmidt, J. O., Roachell, W., Hypša, V., & **Novakova, E.** (2018). Microbiomes of North American Triatominae: the grounds for Chagas disease epidemiology. *Frontiers in Microbiology*, *9*, 1167.

Book Chapters:

1. Hypša, V., & **Nováková, E.** (2009). Insect symbionts and molecular phylogenetics. *Insect symbiosis*, *3*, 1-31.
2. Wilkes, T. E., Duron, O., Darby, A. C., Hypša, V., **Nováková, E.,** & Hurst, G. D. (2011). 12 The Genus *Arsenophonus*. *Manipulative tenants: bacteria associated with arthropods*, 225.

ATTACHED PUBLICATIONS

Paper 1:

Nováková, E., Hypša, V., Klein, J., Foottit, R. G., von Dohlen, C. D., & Moran, N. A. (2013). Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular Phylogenetics and Evolution*, *68*(1), 42-54.

Molecular Phylogenetics and Evolution 68 (2013) 42–54

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*

Eva Nováková [a,b,*], Václav Hypša [a], Joanne Klein [b], Robert G. Foottit [c], Carol D. von Dohlen [d], Nancy A. Moran [b]

[a] Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre, ASCR, v.v.i., Branisovska 31, 37005 Ceske Budejovice, Czech Republic
[b] Department of Ecology and Evolutionary Biology, Yale University, 300 Heffernan Dr., West Haven, CT 06516-4150, USA
[c] Agriculture & Agri-Food Canada, Canadian National Collection of Insects, K.W. Neatby Bldg., 960 Carling Ave. Ottawa, Ontario, Canada K1A 0C6
[d] Department of Biology, Utah State University, UMC 5305, Logan, UT 84322-5305, USA

ARTICLE INFO

Article history:
Received 21 August 2012
Revised 7 March 2013
Accepted 13 March 2013
Available online 29 March 2013

Keywords:
Aphid
Evolution
Buchnera
Phylogeny
Informative markers

ABSTRACT

Reliable phylogenetic reconstruction, as a framework for evolutionary inference, may be difficult to achieve in some groups of organisms. Particularly for lineages that experienced rapid diversification, lack of sufficient information may lead to inconsistent and unstable results and a low degree of resolution. Coincidentally, such rapidly diversifying taxa are often among the biologically most interesting groups. Aphids provide such an example. Due to rapid adaptive diversification, they feature variability in many interesting biological traits, but consequently they are also a challenging group in which to resolve phylogeny. Particularly within the family Aphididae, many interesting evolutionary questions remain unanswered due to phylogenetic uncertainties. In this study, we show that molecular data derived from the symbiotic bacteria of the genus *Buchnera* can provide a more powerful tool than the aphid-derived sequences. We analyze 255 *Buchnera* gene sequences from 70 host aphid species and compare the resulting trees to the phylogenies previously retrieved from aphid sequences, only. We find that the host and symbiont data do not conflict for any major phylogenetic conclusions. Also, we demonstrate that the symbiont-derived phylogenies support some previously questionable relationships and provide new insights into aphid phylogeny and evolution.

1. Introduction

Aphids form a distinctive insect clade that features considerable variability in interesting biological traits, such as the presence of many distinct, yet genetically identical, forms of females during the life cycle (polyphenism), alternation of sexual and asexual reproduction, sterile soldier morphs, and seasonal alternation between unrelated groups of host plants. These traits vary among species, reflecting a long evolutionary history of biogeographical expansions and contractions and codiversification with plant hosts. Understanding the evolution of these traits thus requires a reliable phylogeny as a framework for particular evolutionary hypotheses. However, except for a few generally accepted aspects, studies on aphid phylogeny have not yet produced a clear picture of many relationships within this group.

Earlier attempts to reconstruct aphid phylogeny using morphology resulted in conflicting evolutionary scenarios (Heie, 1987; Wojciechowski, 1992). While an overall picture describing three distinct groups with either viviparous (Aphididae) or oviparous (Adelgidae and Phylloxeridae) parthenogenetic females was generally accepted, the phylogeny of the most diverse group, the Aphididae, remained unclear. More recent efforts have applied DNA sequence data to the reconstruction of Aphididae phylogeny. An emergent problem with such data is a lack of sufficient phylogenetic signal, mostly ascribed to a rapid adaptive radiation within Aphididae (von Dohlen and Moran, 2000; Martínez-Torres et al., 2001). Recently, this difficulty has been partly compensated by combining data from nuclear and mitochondrial genes (Ortiz-Rivas and Martínez-Torres, 2010); however, even this extended source of information leaves many relationships within Aphididae unresolved, and several evolutionary questions unanswered. Some questions concern the number of origins of particular aspects of biology. For example, three tribes with similar life cycles that include dwarf sexual forms lacking mouthparts (Eriosomatini, Fordini, Pemphigini) have been traditionally classified together in the Eriosomatinae (formerly Pemphiginae) (Heie, 1980; Remaudière and Remaudière, 1997). Furthermore, mainly on the basis of shared

* Corresponding author at: Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre, ASCR, v.v.i., Branisovska 31, 37005 Ceske Budejovice, Czech Republic. Fax: +420 385 310 388.
E-mail addresses: novaeva@paru.cas.cz (E. Nováková), vacatko@paru.cas.cz (V. Hypša), Robert.Foottit@agr.gc.ca (R.G. Foottit), carol.vondohlen@usu.edu (C.D. von Dohlen), nancy.moran@yale.edu (N.A. Moran).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.03.016

life cycle characteristics, Eriosomatinae, Hormaphidinae, and Anoeciinae have been hypothesized to share a most recent common ancestor (Heie, 1987). Most molecular phylogenies fail to support these ideas, however, because they recover Eriosomatinae as paraphyletic, and position Eriosomatinae, Hormaphidinae, and Anoeciinae as unrelated lineages. Other biological questions that could be illuminated by improved phylogenetic reconstruction concern whether the common ancestor of extant Aphididae fed on gymnosperms or angiosperms, and whether transitions between these major plant groups were accompanied by changes in diversification rates.

Inclusion of additional markers with appropriate information capacity, as well as additional taxa to bisect long terminal branches, are useful approaches to resolving ambiguous nodes of a rapid radiation (Whitfield and Lockhart, 2007). Nuclear gene markers would be the obvious, potentially informative additions for reconstructing the late Cretacous radiation of Aphididae. However, due to the possibility of paralogy, designation of orthologous nuclear sequences across a broad selection of species could require extensive experimental work. This is particularly relevant for aphids, because a vast amount of gene duplication affecting more than 2000 gene families was confirmed within the *Acyrthosiphon pisum* genome (The International Aphid Genomics Consortium, 2010). This extensive duplication of nuclear genes has been ongoing during aphid evolution, resulting in complex gene trees in which orthologs and paralogs are difficult to distinguish and the species phylogeny is obscured (e.g., Rispe et al., 2008; Nováková and Moran, 2012).

Aphids fortunately possess an additional source of inherited genetic material apart from their own genomes, in the form of obligate, maternally transferred, and highly derived bacterial mutualists. The association between aphids and bacteria of the genus *Buchnera* is one of the most extensively studied symbiotic systems since Buchner (1953) suggested their mutualistic association. As in many other mutualistic relationships, the bacteria supply essential nutrients to their hosts, which in turn provide a stable environment for their bacterial partners. Genomic studies have revealed that particular amino acids, vitamins and sterols are supplemented by *Buchnera* (Akman Gunduz and Douglas, 2009; Moran et al., 2005; Moran and Degnan, 2006; Moya et al., 2008; Hansen and Moran, 2011). *Buchnera* is one of numerous groups of insect-associated symbionts for which evidence supports a long-term, strict cospeciation with the host (Moran et al., 1993, 1995; Clark et al., 2000). The acquisition of *Buchnera* symbionts by aphids is thought to have been a single event that took place by 150–200 MYA and was followed by cospeciation (Moran et al., 1993; Martínez-Torres et al., 2001). Due to this close evolutionary association, the symbiont and host phylogenies mirror each other for deeper evolutionary divergences. Thus, symbiont-derived data in principle can be used to reconstruct the evolutionary history of hosts.

Such an approach is appealing, because the small and simple genomes of symbionts may be an easier source of suitable sequences for phylogenetic analysis. This idea has been considered for some other insect-symbiont associations. For example, Kölsch and Pedersen (2009) suggested using endosymbionts for elucidating unresolved questions of reed beetle phylogeny. In addition to resolving host relationships, endosymbiont genes have been used for calculating divergence times in their hosts (e.g., *Cryptocercus* woodroaches (Maekawa et al., 2005)). *Buchnera* genomes consist of single-copy genes, most of which are shared across different *Buchnera* genomes (Moran et al., 2009); thus, they lack the complications presented by gene duplication and paralogy. Bacterial genomes possess several additional advantages, such as haploidy and absence of introns. Attempts to infer aphid and *Buchnera* phylogenies in a common framework have been undertaken several times (e.g., Munson et al., 1991; Moran et al., 1993; Martínez-Torres et al., 2001), mainly with the aim of testing phylogenetic correspondence of host and symbiont. Only a few studies were focused explicitly on reconstructing aphid phylogeny using *Buchnera* markers; these included analyses performed on *Uroleucon*, *Brachycaudus*, and *Mollitrichosiphum* species (Clark et al., 2000; Jousselin et al., 2009; Liu et al., 2013). All three studies ruled out the possibility of occasional horizontal transfers, even between closely related aphids sharing the same host plant. Even within species, phylogenies based on *Buchnera* genes are congruent with those based on mitochondrial sequences, confirming that this symbiont is strictly maternally inherited (Peccoud et al., 2009; Funk et al., 2000). Within an individual aphid, *Buchnera* genes also have higher copy number than aphid nuclear genes, so that *Buchnera* genes are relatively easy to amplify. DNA of *Buchnera* symbionts thus serves as a useful source of information for inferring aphid phylogeny. However, phylogenetic studies using *Buchnera* genes so far have been limited by sparse sampling across aphid taxa, and especially by taxonomic bias towards the subfamily Aphidinae.

In this study, we use *Buchnera*-derived sequences of five genes (*groEL*, *trpB*, *dnaB*, *ilvD* and 16S rDNA) from an extended taxonomic set to reconstruct phylogenetic relationships within the Aphididae. We then compare the resulting topologies to those reported from recent multilocus analyses of aphid-derived genes. We focus mainly on the most problematic questions in aphid evolution, such as monophyly of individual subfamilies/tribes and their relationships, variation in DNA substitution rates, and rooting of the Aphididae tree. We show that symbiont genes yield informative phylogenetic signal and have several methodological advantages.

2. Materials and methods

2.1. Sample collection

Our study was designed to reconstruct evolutionary relationships among a wide diversity of aphids, represented by a broad sample of most major Aphididae taxa. The collection includes 70 species from 15 of 25 subfamilies and 25 of 36 tribes recognized in the most recent, comprehensive classifications of aphids (Remaudière and Remaudière, 1997; Nieto Nafría et al., 1997) (Table S1).

2.2. DNA extraction, primer design and sequencing

For all species, several individuals were pooled and homogenized, and genomic DNA was extracted using QIAamp DNA Micro Kit (Qiagen). In efforts to compile sufficient data for a reliable phylogenetic reconstruction, we aimed to amplify five *Buchnera*-derived genes, namely 16S rDNA, *groEL*, *trpB*, *dnaB*, and *ilvD*. To confirm the aphid host species identification, the sequence of the aphid mitochondrial gene COI was obtained for each sample.

Three primer pairs allowing for nine combinations were designed for each of the targeted *Buchnera* genes, except for 16S rDNA. Pairs of highly degenerate primers, corresponding to the coding region of *groEL*, *trpB*, and *dnaB* genes, were designed according to alignments of partial sequences available in GenBank using Primaclade software (Gadberry et al., 2005). Primers for *ilvD* were designed in CLC Genomic Workbench (CLC bio A/S) based on an alignment of *Buchnera* genome sequences, with the reverse primers corresponding to a highly conserved region of the flanking tRNA. Primers for amplifying the 16S rDNA gene and aphid mitochondrial gene COI were from previous studies (Folmer et al., 1994; Hajibabaei et al., 2006; Munson et al., 1991; Mateos et al., 2006). More detailed information on primer pairs, including the sequence and the length of amplified regions, is summarized in Table S2.

To amplify PCR products from diverse aphid lineages, we used the two best-performing primer pairs for each gene (Table S2). In

the case of the 16S rRNA gene, primers 10F and 35R were used to detect the presence of secondary symbionts commonly associated with aphids (Fukatsu and Ishikawa, 1993; Sandström et al., 2001; Russell et al., 2003). This primer pair spans roughly a 1600 bp region of the 16S-23S rRNA operon in most bacterial genomes. However, in the reduced genomes of *Buchnera*, genes encoding 16S rRNA and 23S rRNA are widely separated on the chromosome (Munson et al., 1993). Thus, in contrast to almost all other bacteria, the 10F/35R primer pair does not produce an amplicon in *Buchnera*. According to results of the first amplification, the product of the reaction using the universal 16S rRNA primer pair (10F and 1507R, amplifying most of the 16S rRNA gene) was directly sequenced or was subcloned to enable separation of the *Buchnera* and secondary symbiont sequences.

To recover *Buchnera* 16S rDNA sequences in samples with positive 10F/35R primed products indicating presence of other bacteria, amplicons of successful 10F/1507R PCRs were subcloned into Promega pGEM-T Easy vectors. On average, three transformant colonies from each species were picked and their inserts amplified by colony PCR using T7 and SP6 primers. Products of colony PCR as well as PCR products for *groEL*, *trpB*, *dnaB*, *ilvD*, and the mitochondrial COI gene were Sanger-sequenced in both directions on an ABI3700 sequencer at the Yale University DNA sequencing service. Resulting reads were assembled into a single contig using CLC Genomic Workbench (CLC bio A/S) and manually curated to remove base-calling errors. Except for two sequences of *dnaB* amplified from Greenideinae species, bidirectional sequencing did not show any variability. Ambigious positions in *dnaB* were replaced with an "N".

2.3. Alignments and phylogenetic analysis

Accession numbers of the sequences obtained in this study and those retrieved from GenBank for the phylogenetic analyses are listed in Table S3. Matrices for individual genes were compiled from the data obtained in this study and from data retrieved from GenBank, including corresponding *Buchnera* sequences and sequences for selected outgroup taxa. Outgroup sequences were chosen from the *Ishikawaella capsulata* genome (Hosokawa et al., 2006), which shows the lowest pairwise divergence from *Buchnera*, as well as other members of the Gammaproteobacteria, including *Serratia symbiotica*, *Salmonella enterica*, and *Escherichia coli.* Complete datasets for *groEL*, *trpB*, *dnaB*, and *ilvD* were aligned using the ClustalW algorithm as implemented in SeaView software (Gouy et al., 2010). 16S rDNA sequences were aligned in server-based program MAFFT, http://mafft.cbrc.jp/alignment/server/index.html, using the Q-INS-i algorithm with default parameters, which allows for a more accurate alignment of variable loops of the rRNA molecule. The raw alignments were manually corrected in SeaView (Gouy et al., 2010) and further processed in GBlocks (Castresana, 2000) to remove unreliably aligned positions.

The resulting matrices were analyzed using maximum likelihood (ML) and Bayesian inference (BI). To account for the possibility of unreconcilable substitutional saturation in complete datasests, all sequences were analyzed using only the first two codon positions versus all three codon positions for DNA-based analyses, and were also analyzed using amino acid translations for protein-coding genes. The evolutionary models best fitting the data, as well as parameters for ML analyses, were selected using programs jModelTest (Posada, 2008) and ProtTest 3 (Darriba et al., 2011). ML analyses and 100 bootstrap replicates were performed for all nucleotide and protein matrices with PhyML 3.0 (Guindon and Gascuel, 2003) with substitution models and parameters estimated from the data. Bayesian analyses were performed with the closest approximation of best-fit evolutionary models (Table 1) as implemented in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) in two independent runs with four chains. The analyses were run for three to twelve million generations with tree sampling every 100 generations. Exploration of MCMC convergence and burn-in determination was based on values of the average standard deviation of split frequencies, potential scale reduction factor (PSRF), and the log likelihood of the cold chain. Burn-in corresponded to 20–25% of the sampled trees. All the presented topologies are majority-rule consensus trees calculated from the post-burn-in tree samples.

The approximately unbiased (AU) test (Shimodaira and Hasegawa, 2001) was used to compare topologies in two ways. First, we used the AU test to determine whether hypothetical alternative topologies were significantly different from those retrieved by our analyses based on individual genes. In particular, we tested the monophyly/paraphyly of the subfamily Eriosomatinae. Second, we evaluated robustness of the phylogenetic signal by applying the AU test to the best ML topology for each gene and a topology with an obvious introduced artifact, i.e., polyphyly of Aphidinae. To obtain the alternative topologies, we constrained particular artifacts and constructed the tree under the ML criterion in PAUP* 4.0 (Swofford, 2003). To optimize computational demands for these analyses, we compiled simplified matrices for each gene consisting of one representative for each cluster consistently found in all ML and BI derived topologies (not shown). Significant differences between likelihood site patterns for each topology pair were tested using CONSEL version 0.1i (Shimodaira and Hasegawa, 2001). The results of the AU tests were used to identify the set of genes with sufficient information for phylogenetic inference (i.e., rejecting the artificial topology). Two concatenated matrices were constructed from the selected genes. The protein concatenate comprised amino acid sequences of the coding regions for *groEL*, *dnaB*, *trpB*, and *ilvD*. The nucleotide concatenate also included 16S rRNA sequences. Both of the matrices were analyzed as described above.

2.4. Tests of heterogeneity in evolutionary rates of symbiont genes

In *Buchnera* genes, rates of evolution at sites under purifying selection, such as nonsynonymous sites in protein-coding genes, are affected by the host population size and by the number of cells transferred from mother to offspring (Rispe and Moran, 2000). In addition to providing new data useful for inferring aphid phylogeny, our *Buchnera* dataset provides the opportunity to assess whether symbiont population structure differs among aphid taxa. Differences among aphid groups in life cycles, ecology, and transmission of symbionts are expected to produce different levels of purifying selection, leading to rate heterogeneity among aphid groups. Specifically, we tested for differences among groups in the rate of evolution at nonsynonymous sites (d*N*) relative to the rate at synonymous sites (d*S*).

We computed d*N*/d*S* for pairs of sister taxa in each lineage. Average d*N*/d*S* values were then compared among subfamilies, and the data were tested for the presence of statistically deviant values using Dixon's Q test (Rorabacher, 1991). Because d*N*/d*S* ratios are meaningless if changes at synonymous sites are saturated, the saturation index for synonymous positions was determined for each dataset, as calculated in DAMBE (Xia and Xie, 2001), with a proportion of invariable positions assessed in JModelTest (Posada, 2008). Both saturation pattern tests and d*N*/d*S* calculations were performed only with the sites available for all analyzed taxa. These analyses were not performed for subfamilies represented by a single species or by a single pair of species.

2.5. Long branch attraction (LBA)

Preliminary results indicated that some relationships may have been affected by LBA, i.e., artifactual clustering of rapidly evolving taxa regardless of their true relationship (for review see Bergsten,

Table 1
Basic characteristics of matrices and specification of the evolutionary models used in ML and BI analyses.

Alignment designation	Number of taxa	Number of characters	Variable characters	Selected model AICc	Mr. Bayes
trpB	60	489	311	GTR + I + Γ	GTR + I + Γ
dnaB	52	1209	677	GTR + I + Γ	GTR + I + Γ
groEL	85	1135	666	GTR + I + Γ	GTR + I + Γ
ilvD	77	801	520	GTR + Γ	GTR + Γ
16S rRNA	65	1494	1087	GTR + Γ	GTR + Γ
COI	71	690	283	TIM2 + Γ	GTR + Γ
trpB_dnaB_groEL_ilvD_16S	94	5079	2700	TIM3 + Γ	GTR + Γ
trpB_dnaB_groEL_ilvD_1st2nd	94	2390	1021	GTR + I + Γ	GTR + I + Γ
trpB_dnaB_groEL_ilvD_AA	94	1195	662	MtREV + I + Γ + F	WAG + I + Γ

2005). We focused on Aphidinae, where clustering of *Aspidophorodon longicaudus*, *Capitophorus hudsonicus* and two *Pterocomma* species would, if confirmed, contradict monophyly of the tribes Macrosiphini and Pterocommatini. We first tested for effects of LBA on positions of *A. longicaudus* and *C. hudsonicus* in the context of the entire tree, particularly their possible affinity to long-branched taxa of long-branched lineages from other subfamilies (e.g., Saltusaphidinae, Calaphidinae). Additional analyses were performed exclusively within Aphidinae between the four taxa mentioned above. For each analysis, a single branch was removed from the Aphidinae data sets.

3. Results

3.1. Datasets

Altogether, this study generated 255 new sequences from 70 *Buchnera* lineages. Characteristics of the resulting alignments are detailed in Table 1. Because individual single-gene matrices differ in taxon sampling, all of the concatenated matrices are partially incomplete. Missing sequences were either not available in GenBank (for taxa from previous studies), or, for our samples, we were not able to amplify the genes due to PCR failures, most likely caused by substitutions in the primer sites, loss of the gene or even possible absence of *Buchnera* in certain samples. For example, a PCR failure apparently related to primer specificity was experienced for combinations of *dnaB* primers, which failed to amplify sequences for all samples from Calaphidinae and Saltusaphidinae subfamilies. Also, only a few representatives of the subfamily Lachninae yielded PCR products for *trpB*.

Individual matrices of the protein-coding genes contained comparable proportions of variable characters, with the largest proportion found in *trpB* and *ilvD* (64% and 65%, respectively; Table 1). AU tests with constrained topologies indicated different strengths of phylogenetic signal for different matrices. Sequences for *trpB*, *groEL*, *ilvD* and the 16S rRNA gene contained sufficient information to reject all of the tested artificial topologies (see Section 2). For the *dnaB* dataset, the test was significant only when a major topological artifact was introduced (data not shown). In contrast, the mitochondrial COI dataset contained few informative characters (Table 1), and the AU tests were repeatedly insignificant. This dataset was therefore not used for phylogenetic inference; COI sequences were used only to verify species determination based on nucleotide similarity with sequences already available in several copies in GenBank. COI sequences recovered in this study shared similarity above 99% with GenBank homologs from the same species, and thus as redundant records were not deposited in the database.

Nucleotide composition, measured as % GC content, varied by 10% across taxa. The lowest GC levels were found in *Buchnera* from *Pseudoregma koshuensis*, which averaged 27.7% for protein-coding genes. A bias towards higher AT content was also observed in *Buchnera* sequences from Drepanosiphinae, Greenideinae and Lachninae. On the contrary, most of the *Buchnera* sequences isolated from Macrosiphini showed a higher GC content, averaging 34.0% for coding genes.

3.2. Phylogenetic analyses of single-gene matrices

For all genes, analyses restricted to first and second codon positions yielded results similar to those obtained from full nucleotide sequences. However, measures of statistical support (i.e. bootstrap and posterior probabilities) as well as phylogenetic resolution became weaker with decreasing number of positions (i.e., complete DNA sequence, two positions only, and translation into amino acids). This result, together with tests based on a few amino acid matrices (results not shown), indicates that amino acid sequences would yield only poorly supported and largely unresolved topologies. Consequently, amino acid sequences were not used in the single-gene matrices.

All topologies derived from the single-gene datasets placed most aphid species into clades corresponding to tribes of the current classification. For higher taxonomic levels, results were less conclusive, and branch support values were weaker. However, monophyly of Aphidinae, Lachninae, Greenideinae and Chaitophorinae was recovered across different analyses and was well supported. The topology inferred from the *ilvD* single-gene matrix (the dataset with the highest proportion of informative sites) is shown in Fig. 1. Within Aphidinae, Aphidini was monophyletic; however, two species of Macrosiphini, *Aspidophorodon longicaudus* and *Capitophorus hudsonicus*, grouped with *Pterocomma* in most of the single-gene analyses. This grouping was recovered repeatedly in all LBA-avoiding tests, performed on a simplified dataset with one of the long branches removed in each analysis (an example topology shown in Fig. S1).

Monophyly of Hormaphidinae was recovered only in ML analysis of the *trpB* dataset. Two tribes, Hormaphidini and Cerataphidini, did not form a monophyletic group in any other single-gene analysis. Sequences derived from *Pseudoregma koshuensis* (Hormaphidinae: Cerataphidini) had the lowest GC content and clustered artifactually with other sequences of markedly biased nucleotide content. For instance, in the *dnaB* based topology, *Pseudoregma* fell within the same cluster as Greenideinae and Lachninae, which displayed the lowest GC content values for *dnaB* (Figs. S2 and S3). The same pattern was found in the *ilvD* dataset, where *Pseudoregma* clustered as a close relative to significantly AT-biased sequences of the subfamily Drepanosiphinae (Fig. 1). Similar to results in Ortiz-Rivas and Martínez-Torres (2010), Hormaphidini and Thelaxinae were supported as sister taxa. This relationship was supported by all analyses based on single genes, except for *ilvD*.

In most single-gene analyses, Calaphidinae was paraphyletic with respect to Saltusaphidinae. The only exception was the topology derived from *trpB* (Fig. S4), in which Saltusaphidinae was not represented and Calaphidinae was monophyletic. In contrast to the results of Ortiz-Rivas and Martínez-Torres (2010), none of

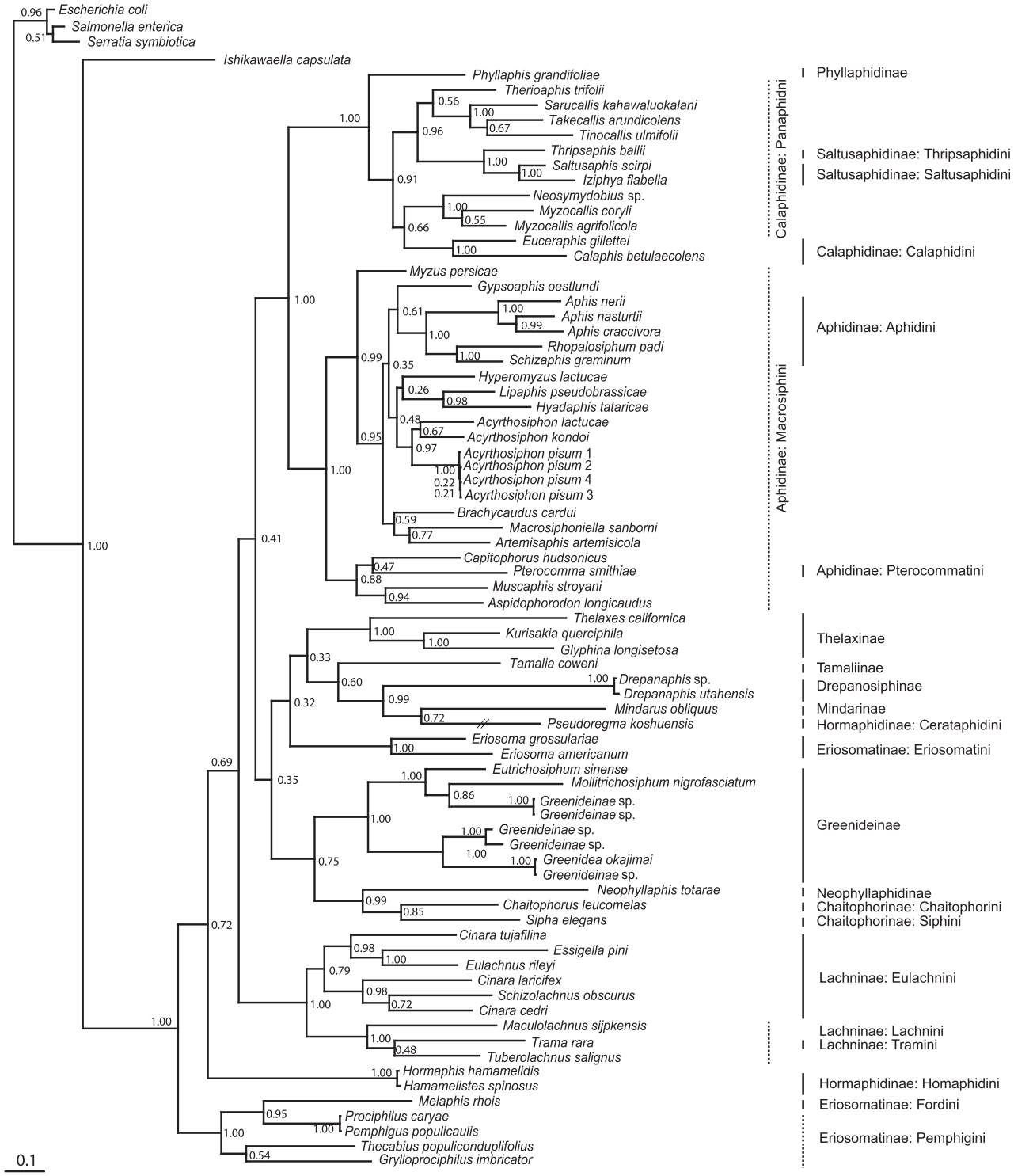


Fig. 1. BI topology based on ilvD dataset. Solid vertical lines refer to monophyletic tribes and subfamilies. Dashed vertical lines designate paraphyletic taxa. The branch of Pseudoregma koshuensis is scaled to ½ of its actual length. The values at nodes show posterior probability support.

our single-gene derived topologies supported monophyly of Drepanosiphidae *sensu* Heie (1980), i.e., including Calaphidinae, Chaitophorinae and Drepanosiphinae.

Eriosomatinae was polyphyletic or paraphyletic in all single-gene analyses. For *ilvD*-derived topologies, Fordini and Pemphigini formed a monophyletic group while Eriosomatini species were placed at a distant position (Fig. 1). A similar result was previously reported for aphid nuclear gene-derived topologies by Ortiz-Rivas and Martínez-Torres (2010). All other single-gene analyses recovered topologies in which the tribes were distant from each other (e.g., Figs. S2 and S4). However, when a constraint of Eriosomatinae monophyly was tested on individual gene sets, all AU test values

were not significant; thus, the monophyly of Eriosomatinae could not be confidently rejected ($p = 0.08–0.47$).

In most of the rooted single-gene topologies, Pemphigini formed the sister group to all other taxa. However, branching of the other deep nodes was inconsistent, varying with datasets and analytical methods, particularly for subfamilies represented by a single specimen, namely *Tamalia coweni* (Tamalinae), *Anoecia oenotherae* (Anoecinae), *Neophyllaphis totarae* (Neophyllaphidinae) and *Mindarus obliquus* (Mindarinae). The overall topological characteristics sorted into three different arrangements. The first supported three distinct lineages: Pemphigini, a monophyletic Aphidinae + Calaphidinae, and a cluster composed of all remaining subfamilies and tribes. In the other two scenarios, aphid diversity was distributed into two major clusters, either Aphidinae versus a lineage of all other taxa, or (Aphidinae + Calaphidinae) versus a lineage of all remaining taxa.

3.3. Phylogenetic analyses of concatenated data sets

In order to test the potential impact of data incompleteness in concatenated matrices, an additional *groEL + trpB* dataset was constructed from the taxa for which sequences of both genes were available. These genes were chosen because they were available for the largest number of taxa. The topology obtained from both BI and ML corresponded to the scenario described above with Aphidinae versus a lineage of all other taxa, and did not differ substantially from the results obtained from analyses of the incomplete concatenated matrices.

Similarly to the single-gene topologies, most results from concatenated datasets placed Pemphigini as sister to the remaining taxa. Two main types of topology were repeatedly retrieved from the concatenated data. While most BI analyses tended to produce topologies with three main clusters, e.g. Pemphigini, Aphidinae + Calaphidinae, and a clade comprising the rest of taxa (Figs. 2 and 3), a few ML analyses supported grouping into two major lineages with one consisting exclusively of the subfamily Aphidinae (not shown).

The topologies recovered in this study are partially congruent with the recently published aphid phylogeny inferred from two nuclear and one mitochondrial genes (Ortiz-Rivas and Martínez-Torres, 2010). The overall similarity between the aphid-derived and *Buchnera*-derived trees provides sufficient background for detailed comparisions and evaluation of particular phylogenetic and evolutionary patterns. Ortiz-Rivas and Martínez-Torres (2010) recovered three major aphid clusters, namely Lachninae, a clade containing Eriosomatinae + Anoeciinae + Thelaxinae + Hormaphidinae + Mindarinae, and a clade containing Aphidinae + Calaphidinae + Drepanosiphinae + Chaitophorinae. Lachninae formed the sister group to a lineage comprising the latter two clades in their results. In contrast, we consistently recovered Pemphigini as the sister group to a lineage including all other taxa (Figs. 2–4). None of our analyses produced a topology in which Lachninae branched from the base of the tree, although this position was not rejected by the AU tests. Furthermore, only the ML analysis based on concatenation of nucleotides for 1st and 2nd positions produced a cluster containing Aphidinae + Calaphidinae + Drepanosiphinae + Chaitophorinae, as in Ortiz-Rivas and Martínez-Torres (2010). In this topology, the cluster also included Saltusaphidinae and Phyllaphidinae, which were not sampled by Ortiz-Rivas and Martínez-Torres (2010). For other recovered topologies, only the close relationship between Aphidinae and Calaphidinae was highly supported. Thus, the third cluster from the Ortiz-Rivas and Martínez-Torres study, Eriosomatinae + Anoeciinae + Thelaxinae + Hormaphidinae + Mindarinae, was never recovered (Figs. 2–4).

Similar to the topologies derived from aphid genes (Ortiz-Rivas and Martínez-Torres, 2010), we obtained only low support for several deep nodes (Fig. 2). Particularly, the position of the subfamilies represented by a single specimen, namely *Tamalia coweni* (Tamalinae), *Anoecia oenotherae* (Anoecinae), and *Neophyllaphis totarae* (Neophyllaphidinae) appeared to be unstable under different analysis settings. Bootstrap support for topologies calculated under the ML criterion was low, possibly due to the incompleteness of concatenated matrices (Fig. 4).

Because the outgroups selected for all analyses are relatively distant to *Buchnera*, unrooted analyses (i.e. analyses excluding outgroups) were also performed with concatenated datasets to avoid potentially misleading information introduced by very distant sequences. An example of topologies consistently retrieved, regardless of method or gene composition of concatenates, is shown in Fig. 5. Taxa in this unrooted trichotomy fell into three main clusters: (1) Lachninae, (2) Neophyllaphidinae, Chaitophorinae, Anoeciinae, Cerataphidini, Drepanosiphinae, Thelaxinae, Greenideinae, and (3) Aphidinae, Calaphidinae, Saltusaphidinae, Phyllaphidinae, Tamalinae, Hormaphidini, Eriosomatinae (paraphyletic). A close relationship between Aphidinae + Calaphidinae is highly supported, as is paraphyly of Hormaphidinae (Hormaphidini and Cerataphidini positioned in two different clusters).

3.4. Tests of heterogeneity in evolutionary rate of symbiont genes

Comparisons of average d*N*/d*S* indicated heterogeneity in evolutionary rates across subfamilies and tribes. Elevated values of d*N*/d*S* were found for Calaphidinae, Chaitophorinae, Greenideinae and Saltusaphidinae (Table 2), all groups with simple life cycles containing few distinct female morphs and narrow host-plant ranges. Rates at nonsynonymous sites were not elevated in the Lachninae, nor in Aphidini, Pemphigini, or Hormaphidinae, which include species with complex, host-alternating life cycles.

4. Discussion

4.1. Phylogenetic inference: comparison of Buchnera and aphid data

The results obtained here show that the *Buchnera* genes selected for this study are useful sources of data for corroboration or rejection of particular phylogenetic hypotheses, although they cannot completely resolve aphid phylogeny. Here we summarize how our results compare to the most recent phylogeny for Aphididae based on aphid genes (Ortiz-Rivas and Martínez-Torres, 2010). First, relationships strongly supported by both *Buchnera* and aphid sequences include monophyly of individual tribes and monophyly of Lachninae. Second, relationships congruent with those from aphid genes but more strongly supported by *Buchnera* sequences include paraphyly of Macrosiphini with respect to Pterocommatini and monophyly of Aphidinae (including Aphidini, Macrosiphini, Pterocommatini). Third, relationships incongruent between analyses based on *Buchnera* versus aphid genes, with neither providing strong or stable resolution, were concentrated at the deeper nodes. An example is the monophyly versus polyphyly of Eriosomatinae: while Fordini was repeatedly recovered as a sister group to Eriosomatini in *Buchnera*- based topologies, the aphid data supported a sister relationship of Fordini and Pemphigini. Other instances of unstable resolution by either aphid or *Buchnera* genes involve relationships between Calaphidinae and Aphidinae, and among the Calaphidinae, Chaitophorinae, and Drepanosiphinae, and relationships between Anoeciinae and Mindarinae. Most notably, we did not find any clear example of strongly supported discrepancies between analyses based on aphid genes versus *Buchnera* genes. We note that differences or lack of differences between the studies also could be affected by the differences in taxon sampling.

Fig. 2. BI topology based on the concatenated matrix including 1st and 2nd codon positions. Thicker lines designate branches with a posterior probability above 95%. Solid vertical lines refer to monophyletic subfamilies. Dashed vertical line designates paraphyletic/polyphyletic taxa.

This overview demonstrates that, despite particular differences among the many trees recovered in this study and in Ortiz-Rivas and Martínez-Torres (2010), both data sources are highly congruent for strongly supported relationships, even though they may differ for weakly supported nodes. This suggests that the inconclusiveness of previous studies based on aphid DNA reflects real evolutionary processes (e.g., rapid diversification in some lineages) rather than methodological artifacts.

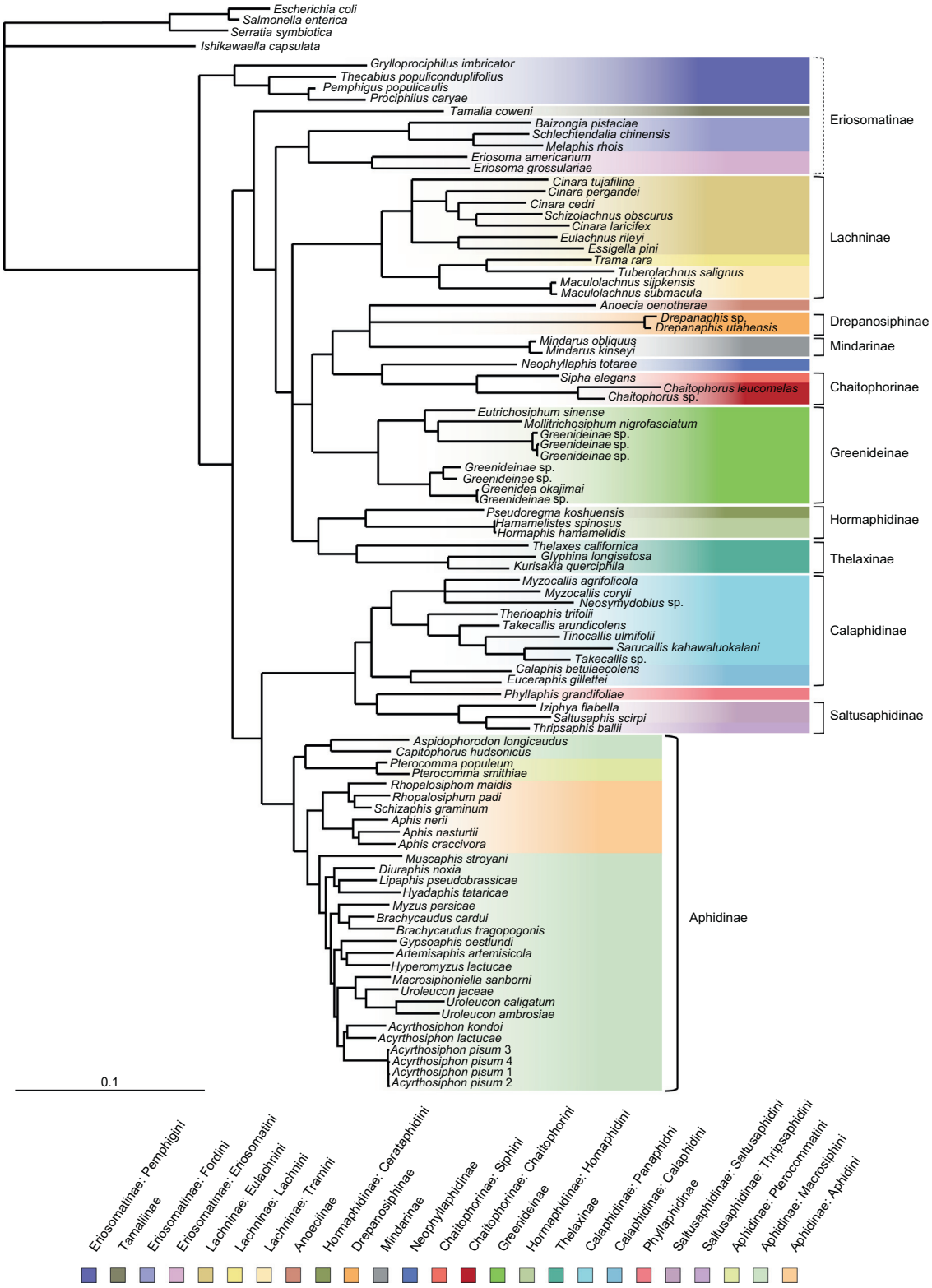


Fig. 3. A combined BI topology constructed to maximize the information content of amino acid and nucleotide data. The overall tree is based on concatenated amino acids, and the relationships within Aphidini, Macrosiphini and Pterocomatini are based on concatenated nucleotide sequences corresponding to the same matrix (including all three codon positions). Solid vertical lines refer to monophyletic subfamilies.

Apart from a general assessment of the suitability of *Buchnera* data for resolving aphid phylogeny, our study also found unexpected relationships, suggesting the need for taxonomic reassessment in some groups. For example, two genera classified as Macrosiphini, *Aspidophorodon longicaudus* and *Capitophorus hudsonicus*, grouped significantly with *Pterocomma* (Pterocommatini)

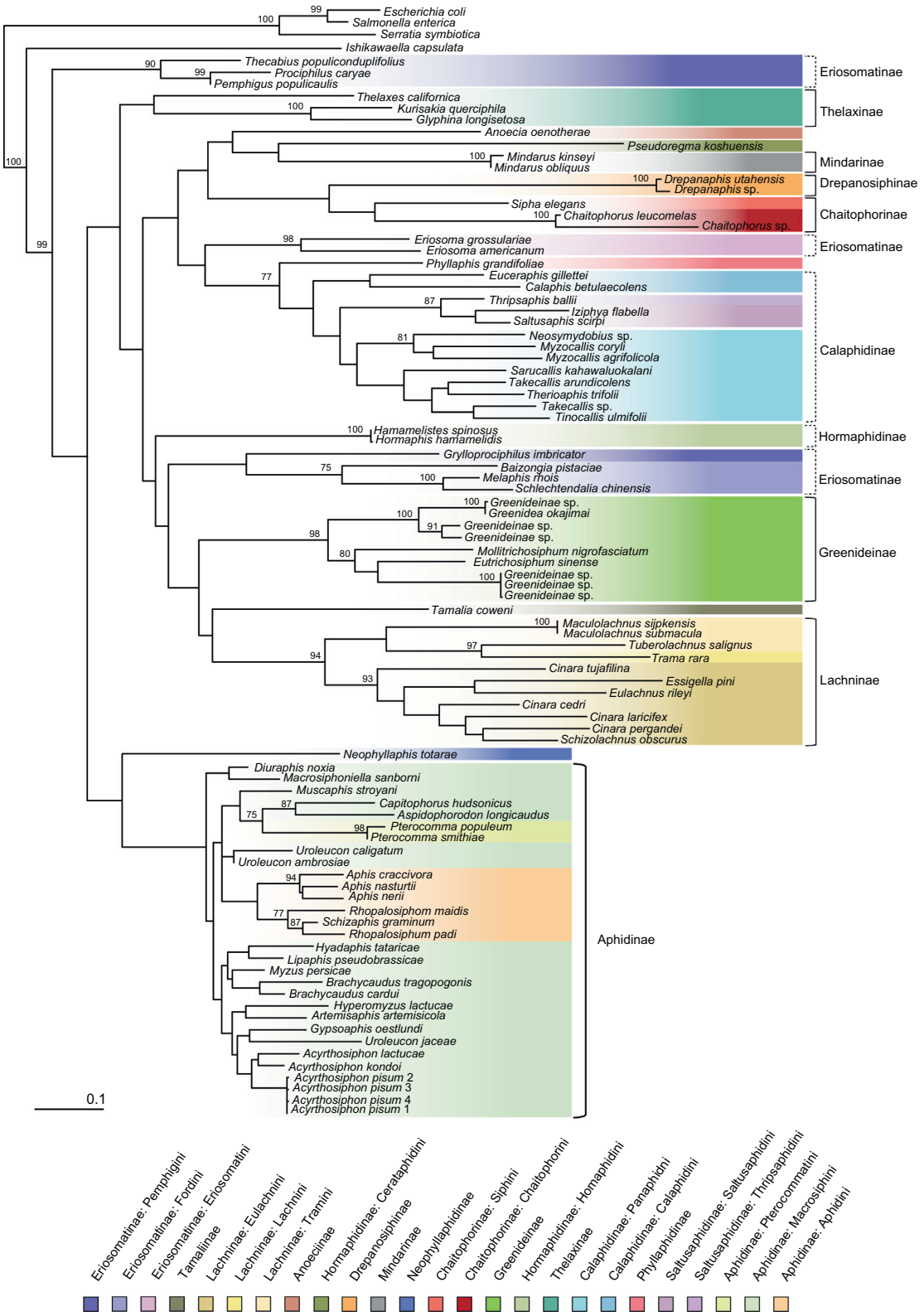


Fig. 4. ML topology based on the concatenated amino acid matrix. Bootstrap support values above 70% are shown. Solid vertical lines refer to monophyletic subfamilies. Dashed vertical lines designate paraphyletic/polyphyletic taxa.

rather than with other Macrosiphini, thus rendering Macrosiphini paraphyletic. Clustering of these three species was supported by analyses of single-gene topologies and by concatenated matrices. Furthermore, a possible effect of LBA on the position of these taxa was tested and discounted, suggesting that their relationship reflects evolutionary history rather than phylogenetic artifact. A

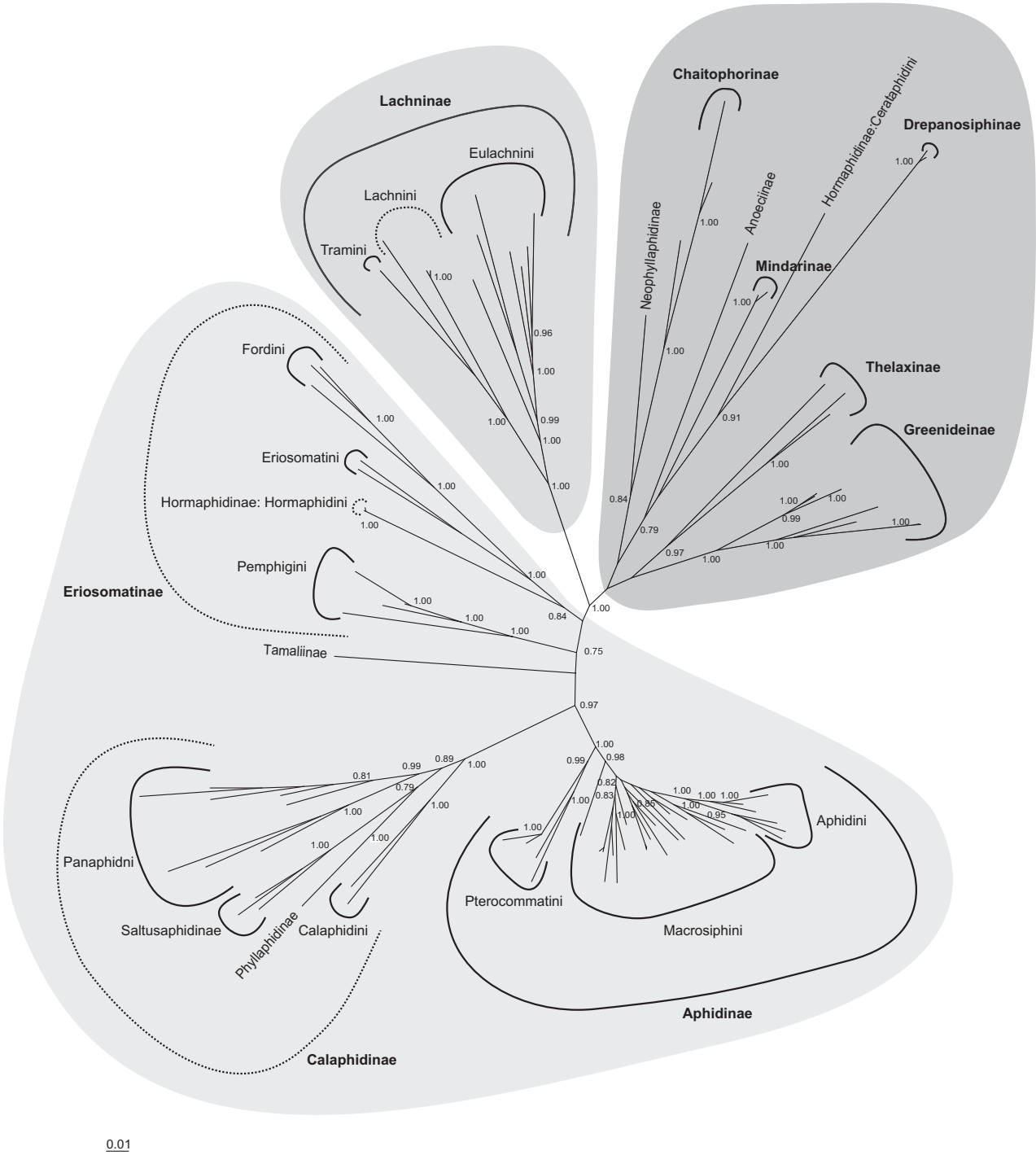


Fig. 5. Unrooted BI topology based on the concatenated matrix including 1st and 2nd codon positions. Three main clusters are defined (shaded areas) to facilitate comparison with the Ortiz-Rivas and Martínez-Torres (2010) scheme, and are resolved as a trichotomy in this unrooted analysis. Solid lines refer to monophyletic taxa. Dashed lines designate paraphyletic/polyphyletic taxa. Numbers are posterior probabilities.

previous study (von Dohlen et al., 2006) also recovered a paraphyletic Macrosiphini, due to the sister relationship of *Pterocomma* with *Cavariella* (a genus not included in the present study); a later study (Kim et al., 2011) duplicated this finding. Together, these studies indicate the need for a re-examination and probable taxonomic revision of Macrosiphini and Pterocommatini.

Some of the relationships found in this study have implications for the evolution of host-plant relationships and other interesting features of aphid biology. One example concerns Lachninae. Normark (2000) found previously that Tramini was nested within Lachnini, rendering the latter tribe paraphyletic. Our analysis recovers the same pattern using *Buchnera* genes, and with high support. Not only do these results indicate that the tribal classification does not reflect phylogeny, they have implications for the interpretation of feeding strategies within Lachninae. Ortiz-Rivas and Martínez-Torres, 2010 proposed that a single transition from gymnosperm hosts to angiosperms may have occured in an ancestor of Lachnini + Tramini, with Eulachnini retaining the ancestral

Table 2
Saturation pattern and evolutionary rates found for major aphid lineages.

Lineage	Iss[a] all sites	Iss.c[b] all sites	Iss[a] resolved sites	Iss.c[b] resolved sites	Iss.c[c]	d*N*/d*S* average
Aphidinae	0.6124	0.8006	0.1488	0.7577	0.4817	0.0646
Calaphidinae	0.7304	0.8112	0.2120	0.7282	0.5740	0.1263[d]
Fordini	1.0924	0.8426	0.1680	0.8368	0.8310	0.0691
Greenideinae	0.7071	0.8305	0.1358	0.8081	0.7079	0.1191[d]
Hormaphidinae	0.5282	0.8426	0.1825	0.8402	0.8391	0.0864
Chaitophorinae	1.1997	0.8426	0.1447	0.8229	0.8133	0.1053[d]
Lachninae	0.6455	0.8267	0.1816	0.7626	0.6145	0.0609
Pemphigini	0.7388	0.8426	0.1902	0.8252	0.8152	0.0558
Saltusaphidinae	1.0317	0.8405	0.1530	0.8067	0.7985	0.1362[d]
Thelaxinae	0.5449	0.8426	0.2386	0.8418	0.8445	0.0886

[a] Index of substitution saturation (Iss) calculated using DAMBE software.
[b] Critical Iss values for calculations including all sites and resolved sites only.
[c] Iss.c[c] stands for critical Iss values calculated for an asymmetrical tree.
[d] Lineages demonstrated to have elevated d*N*/d*S*.

state of gymnosperm feeding. However, they acknowledged that this scenario was not fully supported by their data, because a lineage comprising Lachnini + Tramini was only rarely present in their trees. The nesting of Tramini within Lachnini, supported strongly by our data, is compatible with the hypothesis of a single transition between gymnosperms and angiosperms within the Lachninae. Another example of a phylogenetic result with implications for the evolutionary history of host plant associations concerns the position of Saltusaphidinae. This subfamily was recovered previously as nested within Calaphidinae (von Dohlen and Moran, 2000); with greater taxon sampling in our study, this nested position was even more strongly supported. This relationship suggests that ancestors of sedge- and rush-feeding Saltusaphidinae shifted directly from feeding on trees (as in extant Calaphidinae) to feeding on herbaceous monocots. In contrast, most other aphid species living on herbaceous plants seem to have acquired these hosts through a host-alternating intermediate life-cycle stage including both trees and herbs (Moran, 1992).

One final example of a phylogenetic result with implications for the evolution of an unusual phenotype relates to the monophyly versus paraphyly of Eriosomatinae, whose members all produce dwarf sexual forms. Two of the three tribes (Eriosomatini and Fordini) are supported as sister groups, with weak support based on the aphid data (Ortiz-Rivas and Martínez-Torres, 2010) and considerably stronger support in our analyses. The third tribe (Pemphigini) does not cluster with Eriosomatini + Fordini. Thus, both *Buchnera* and aphid data suggest that dwarf non-feeding sexuals may have originated twice, through convergent adaptation to the host plant-alternating life style. A more thorough understanding of the evolution of this puzzling morph (i.e. multiple origins versus repeated losses) could be investigated further with more phylogenetic studies and/or developmental data. While several other topological aspects presented here may influence interpretation of the evolution of host-plant associations or life cycles if proven correct, the resolution and support of the deeper nodes are too weak for definite conclusions.

4.2. Rooting the Buchnera tree

The phylogenies derived from *Buchnera* sequences and aphid sequences appear to contradict one another regarding placement of the root node. In analyses based on aphid genes, the root was placed between Lachninae and all other taxa in all presented trees (Ortiz-Rivas and Martínez-Torres, 2010). In contrast, in the *Buchnera*-derived topologies, the root is positioned between Pemphigini and all other taxa in most cases. However, the AU tests indicate that this root position cannot be considered as the only possibility. AU tests were not able to reject a root located between any of the four proposed clusters (including between Lachninae and other taxa). This incongruence and uncertainty may reflect the different taxon sampling and outgroups used in the two studies. While for the *Buchnera*-based analyses the best available outgroup was the phylogenetically distant *Ishikawaella capsulata* symbiont, the aphid-based reconstructions could be rooted with more closely related taxa, e.g., *Daktulosphaira* in Phylloxeridae. However, it is clear that neither study reliably identifies the root.

4.3. Tests of heterogeneity in evolutionary rate of symbiont genes

The d*N*/d*S* calculations reveal higher values for *Buchnera* of aphid groups with simple life cycles and restricted to living on trees, except for Saltusaphidinae, which lives on sedges but is derived from within tree-feeding Calaphidinae. These aphid groups are often considered to represent the ancestral aphid life cycle, involving fewer distinct female morphs and only a single host-plant group. Higher d*N*/d*S* due to reduced efficacy of purifying selection is expected when the aphid population size is smaller or when the number of symbiont cells inoculating each progeny is smaller (Rispe and Moran, 2000). A previous phylogenetic study based on aphid genes (Ortiz-Rivas and Martínez-Torres, 2010) presented analyses of rate differences using relative-rate tests as implemented in RRTree (Robinson-Rechavi and Huchon, 2000). Those results cannot be meaningfully compared to the d*N*/d*S* results for *Buchnera*, as relative rate tests are designed to detect changes in rates between a test lineage and related lineages, whereas d*N*/d*S* comparisons reveal changes in efficacy of purifying selection on different branches. We do not report RRTree analyses because the method is prone to several errors, particularly when the closest known outgroup is distant (Bromham et al., 2000; Robinson et al., 1998), as is the case for *I. capsulat*a (Hosokawa et al., 2006).

5. Conclusion

The results obtained from *Buchnera*-derived sequences confirm the overall suitability of this source of phylogenetic information, and further suggest that *Buchnera* genes may outperform host genes for solving particular phylogenetic problems. A potential drawback that must be considered when using these markers is the well-known tendency of symbiont sequences to suffer from various phylogenetic artifacts, mostly due to LBA and compositional bias. However, in many animal groups, these issues also affect mitochondrial and nuclear sequences; in aphids in particular, most nuclear genes are affected by extensive paralogy. Fortunately, new computational techniques are in development that can avoid artifacts due to LBA and compositional bias, provided that suffi-

cient amounts of data are available (Lartillot et al., 2009; Husník et al., 2011).

Codiversification with obligate bacterial symbionts has now been documented for many invertebrate taxa, including a wide range of insects such as cockroaches, ants, tsetse flies, weevils, lice, and almost all insect groups feeding on plant sap (Allen et al., 2007; Conord et al., 2008; Moran et al., 2008), as well as a diversity of other animal clades such as marine flatworms and some marine clams (Gruber-Vodicka et al., 2011; Peek et al., 1998). Considering the rapidly expanding bacterial genome databases, symbiont-derived data are likely to become a powerful tool for reconstructing host phylogenies in these strictly cospeciating associations.

Acknowledgments

We thank S. Aoki, T. Fukatsu, U. Kurosu, J. Ortego, D. A. J. Teulon and E. Maw for donating samples. This work was supported by the United States National Science Foundation (DEB-0723472, DEB-1106195 to N.A.M.). E.N. was supported on a Fulbright fellowship and by the Grant Agency of the University of South Bohemia (135/2010/P) and the Grant Agency of the Czech Republic (206/09/H026). C.D.v.D. was supported by the Utah Agricultural Experiment Station; approved as journal paper #8462.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.03.016.

References

Akman Gunduz, E., Douglas, A.E., 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proc. Biol. Sci. 276, 987–991.
Allen, J.M., Reed, D.L., Perotti, M.A., Braig, H.R., 2007. Evolutionary relationships of "Candidatus Riesia spp.", endosymbiotic enterobacteriaceae living within hematophagous primate lice. Appl. Environ. Microbiol. 73, 1659–1664.
Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.
Bromham, L., Penny, D., Rambaut, A., Hendy, M.D., 2000. The power of relative rates tests depends on the data. J. Mol. Evol. 50, 296–301.
Buchner, P., 1953. Endosymbiose der Tiere mit pflanzlichen Mikroorganismen. Birkhäuser, Basel, Stuttgart, 771pp.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Clark, M.A., Moran, N.A., Baumann, P., Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution 54, 517–525.
Conord, C., Despres, L., Vallier, A., Balmand, S., Miquel, C., Zundel, S., Lemperiere, G., Heddi, A., 2008. Long-term evolutionary stability of bacterial endosymbiosis in Curculionoidea: additional evidence of symbiont replacement in the Dryophthoridae family. Mol. Biol. Evol. 25, 859–868.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.
Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–297.
Fukatsu, T., Ishikawa, H., 1993. Occurrence of Chaperonin 60 and Chaperonin 10 in primary and secondary bacterial symbionts of aphids: Implications for the evolution of an endosymbiotic system in aphids. J. Mol. Evol. 36, 568–577.
Funk, D.J., Helbling, L., Wernegreen, J.J., Moran, N.A., 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. Proc. Biol. Sci. 267, 2517–2521.
Gadberry, M.D., Malcomber, S.T., Doust, A.N., Kellogg, E.A., 2005. Primaclade - a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21, 1263–1264.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.
Gruber-Vodicka, H.R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N.R., Horn, M., Lott, C., Loy, A., Wagner, M., Ott, J., 2011. *Paracatenula*, an ancient symbiosis between thiotrophic Alphaproteobacteria and catenulid flatworms. Proc. Natl. Acad. Sci. USA 108, 12078–12083.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. USA 103, 968–971.
Hansen, A.K., Moran, N.A., 2011. Aphid genome expression reveals host symbiont cooperation in the production of amino acids. Proc. Natl. Acad. Sci. USA 108, 2849–2854.
Heie, O.E., 1980. The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. I. General Part. The families Mindaridae, Hormaphididae, Thelaxidae, Anoeciidae, and Pemphigidae. Scandinavian Science Press Ltd., Klampenborg, Denmark.
Heie, O.E., 1987. Palaeontology and phylogeny. In: Minks, A.K., Harrewijn, P. (Eds.), Aphids: Their biology, Natural Enemies, and Control, vol. 2A. Elsevier, Amsterdam.
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.
Husník, F., Chrudimský, T., Hypša, V., 2011. Multiple origin of endosymbiosis within the Enterobacteriaceae (gamma-Proteobacteria): convergency of complex phylogenetic approaches. BMC Biol. 9, 87.
Jousselin, E., Desdevises, Y., D'Acier, A.C., 2009. Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. Proc. R. Soc. Biol. Sci. 276, 187–196.
Kim, H., Lee, S., Jang, Y., 2011. Macroevolutionary patterns in the aphidini aphids (Hemiptera: Aphididae): diversification, host association, and biogeographic origins. PLoS ONE 6, e24749.
Kölsch, G., Pedersen, B.V., 2009. Can the tight co-speciation between reed beetles (Col., Crysomelidae, Donaciinae) and their bacterial endosymbionts providing cocoon material clarify the deeper phylogeny of the hosts? Mol. Phylogen. Evol. 54, 810–821.
Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288.
Liu, L., Huang, X.L., Zhang, R.L., Jiang, L.Y., Qiao, G.X., 2013. Phylogenetic congruence between *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect-bacteria parallel evolution. Syst. Entomol. 38, 81–92.
Maekawa, K., Park, Y.C., Lo, N., 2005. Phylogeny of endosymbiont bacteria harbored by the woodroach *Cryptocercus* spp. (Cryptocercidae: Blattaria): molecular clock evidence for a late cretaceous-early tertiary split of Asian and American lineages. Mol. Phylogenet. Evol. 36, 728–733.
Martínez-Torres, D., Buades, C., Latorre, A., Moya, A., 2001. Molecular systematics of aphids and their primary endosymbionts. Mol. Phylogenet. Evol. 20, 437–449.
Mateos, M., Catrezana, S.J., Nankivell, B.J., Estes, A., Markow, T.A., Moran, N.A., 2006. Heritable endosymbionts of *Drosophila*. Genetics 174, 363–376.
Moran, N.A., 1992. The evolution of life cycles in aphids. Ann. Rev. Entomol. 37, 321–348.
Moran, N.A., Degnan, P.H., 2006. Functional genomics of *Buchnera* and the ecology of aphid hosts. Mol. Ecol. 15, 1251–1261.
Moran, N.A., Munson, M.A., Baumann, P., Ishikawa, H., 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B 253, 167–171.
Moran, N.A., von Dohlen, C.D., Baumann, P., 1995. Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. J. Mol. Evol. 41, 727–731.
Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E., Ochman, H., 2005. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. Proc. Natl. Acad. Sci. USA 102, 16919–16926.
Moran, N.A., McCutcheon, J.P., Nakabachi, A., 2008. Evolution and genomics of heritable bacterial symbionts. Ann. Rev. Genet. 42, 165–190.
Moran, N.A., McLaughlin, H.J., Sorek, R., 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323, 379–382.
Moya, A., Pereto, J., Gil, R., Latorre, A., 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat. Rev. Gen. 9, 218–229.
Munson, M.A., Baumann, P., Clark, M.A., Baumann, L., Moran, N.A., Voegtlin, D.J., Campbell, B.C., 1991. Evidence for the establishment of aphid eubacterium endosymbiosis in an ancestor of four aphid families. J. Bacteriol. 173, 6321–6324.
Munson, M.A., Baumann, L., Baumann, P., 1993. *Buchnera aphidicola* (a prokaryotic endosymbiont of aphids) contains a putative 16S rRNA operon unlinked to the 23S rRNA-encoding gene: sequence determination and promoter and terminator analysis. Gene 137, 171–178.
Nieto Nafría, J.M., Mier Durante, M.P., Remaudière, G., 1997. Les noms des taxa du groupe-famille chez les Aphididae (Hemiptera). Revue Fr. Ent. (NS) 19, 77–92.
Normark, B.B., 2000. Molecular systematics and evolution of the aphid family Lachnidae. Mol. Phylogenet. Evol. 14, 131–140.
Nováková, E., Moran, N.A., 2012. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. Mol. Biol. Evol. 29, 313–323.
Ortiz-Rivas, B., Martínez-Torres, D., 2010. Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae. Mol. Phylogenet. Evol. 55, 305–317.
Peccoud, J., Simon, J.C., McLaughlin, H.J., Moran, N.A., 2009. Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts. Proc. Natl. Acad. Sci. USA 106, 16315–16320.
Peek, A.S., Feldman, R.A., Lutz, R.A., Vrijenhoek, R.C., 1998. Cospeciation of chemoautotrophic bacteria and deep sea clams. Proc. Natl. Acad. Sci. USA 95, 9962–9966.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Remaudière, G., Remaudière, M., 1997. Catalogue des Aphididae du Monde. INRA, Homoptera Aphidoidea. Paris, 473pp.
Rispe, C., Moran, N.A., 2000. Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. Amer. Nat. 156, 425–441.
Rispe, C., Kutsukake, M., Doublet, V., Hudaverdian, S., Legeai, F., Simon, J.C., Tagu, D., Fukatsu, T., 2008. Large gene family expansion and variable selective pressures for cathepsin B in aphids. Mol. Biol. Evol 25, 5–17.
Robinson, M., Gouy, M., Gautier, C., Mouchiroud, D., 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. 15, 1091–1098.
Robinson-Rechavi, M., Huchon, D., 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16, 296–297.
Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Rorabacher, D.B., 1991. Statistical treatment for rejection of deviant values: critical values of Dixon Q parameter and related subrange ratios at the 95 percent confidence level. Anal. Chem. 63, 139–146.
Russell, J.A., Latorre, A., Sabater-Muñoz, B., Moya, A., Moran, N.A., 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol. Ecol. 12, 1061–1075.
Sandström, J.P., Russell, J.A., White, J.P., Moran, N.A., 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. Mol. Ecol. 10, 217–228.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
The International Aphid Genomics Consortium, 2010. Genome Sequence of the Pea Aphid Acyrthosiphon pisum. PLoS Biol. 8, e1000313.
von Dohlen, C.D., Moran, N.A., 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. Biol. J. Linn. Soc. 71, 689–717.
von Dohlen, C.D., Rowe, C.A., Heie, O.E., 2006. A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. Mol. Phylogenet. Evol. 38, 316–329.
Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22, 258–265.
Wojciechowski, W., 1992. Studies on the Systematic System of Aphids (Homoptera, Aphidinea). Uniwersytet Slaski, Katowice.
Xia, X., Xie, Z., 2001. DAMBE: data analysis in molecular biology and evolution. J. Heredity 92, 371–373.

Paper 2:

Nováková, E., & Moran, N. A. (2012). Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Molecular biology and evolution*, *29*(1), 313-323.

Diversification of Genes for Carotenoid Biosynthesis in Aphids following an Ancient Transfer from a Fungus

Eva Nováková†,[1] and Nancy A. Moran*[1]

[1]Department of Ecology and Evolutionary Biology, Yale University

†Present address: Institute of Parasitology, Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic.

***Corresponding author:** E-mail: nancy.moran@yale.edu.

Associate editor: Jennifer Wernegreen

Abstract

The pea aphid genome was recently found to harbor genes for carotenoid biosynthesis, reflecting an ancestral transfer from a fungus. To explore the evolution of the carotene desaturase gene family within aphids, sequences were retrieved from a set of 34 aphid species representing numerous deeply diverging lineages of aphids and analyzed together with fungal sequences retrieved from databases. All aphids have at least one copy of this gene and some aphid species have up to seven, whereas fungal genomes consistently have a single copy. The closest relatives of aphids, adelgids, also have carotene desaturase; these sequences are most closely related to those from aphids, supporting a shared origin from a fungal to insect transfer predating the divergence of adelgids and aphids. Likewise, all aphids, and adelgids, have carotenoid profiles that are consistent with their biosynthesis using the acquired genes of fungal origin rather than derivation from food plants. The carotene desaturase was acquired from a fungal species outside of Ascomycota or Basidiomycota and closest to Mucoromycotina among sequences available in databases. In aphids, an ongoing pattern of gene duplication is indicated by the presence of both anciently and recently diverged paralogs within genomes and by the presence of a high frequency of pseudogenes that appear to be recently inactivated. Recombination among paralogs is evident, making analyses of patterns of selection difficult, but tests of selection for a nonrecombining region indicates that duplications tend to be followed by bouts of positive selection. Species of Macrosiphini, which often show color polymorphisms, typically have a larger number of desaturase copies relative to other species sampled in the study. These results indicate that aphid evolution has been accompanied by ongoing evolution of carotenogenic genes, which have undergone duplication, recombination, and occasional positive selection to yield a wide variety of carotenoid profiles in different aphid species.

Key words: carotenoid, lateral gene transfer, gene duplication, insects, aphids, *Acyrthosiphon pisum*.

Research article

Introduction

Carotenoids form a large class of compounds that are present in many organisms and that can confer a variety of benefits, including protection from oxidative damage, light detection, photoprotection, and display coloration (Britton et al. 2004, 2006). Although carotenoids are diverse and ubiquitous in organisms, their production depends on a limited set of enzymes that likely originated early in the evolution of cellular life (Klassen 2010). These include phytoene synthase (CrtB family, which produces the initial 40-carbon backbone of the carotenoid hydrocarbon), carotene desaturase (CrtI family, which introduces double bonds between neighboring carbon residues), and carotene cyclase (CrtL family, which introduces six-carbon rings at the termini of the C40 backbone). Additional enzymes, present in particular groups of organisms, can introduce further modifications, such as the addition of hydroxyl groups or shortening or lengthening of the carbon backbone. The core enzymatic machinery for carotenoid production is encoded in the genomes of many Bacteria, Archaea, unicellular eukaryotes, fungi, and plants. Phylogenetic analyses indicate an ancient origin of carotenoid biosynthesis, followed by limited amounts of horizontal transfer and gene duplication in particular lineages of prokaryotes (Klassen 2010).

Animals require carotenoids, for vision, display coloration, and other cellular functions (Britton et al. 2006), but most animals lack the genes encoding carotenoid biosynthetic machinery and thus must obtain carotenoids from food. Unexpectedly, *Acyrthosiphon pisum* (pea aphid) possesses functional carotenoid biosynthetic genes, providing the first instance of carotenoid production by animals, as well as an unusual case of an animal genome acquiring foreign genes of known function (Moran and Jarvik 2010). Phylogenetic analyses revealed that aphid carotenogenic genes are derived from an ancestral transfer of DNA from a fungal genome to an insect genome.

In both *A. pisum* and certain fungi, phytoene synthase and carotene cyclase are fused and are encoded in the same chromosomal region as carotene desaturase, in a distinctive bidirectionally transcribed arrangement not known from other carotenogenic organisms. In *A. pisum*, this unit has undergone several duplication events subsequent to its acquisition, resulting in four copies of the genes for carotene desaturase and three for the fused phytoene synthase/carotene cyclase. Such duplication of carotenoid

biosynthetic genes is unusual in bacteria, fungi, or plants and potentially reflects subfunctionalization (e.g., Gallagher et al. 2004). Indeed, in *A. pisum*, only one of the four copies of the desaturase is responsible for the production of the red carotenoid torulene (Moran and Jarvik 2010). This observation raises the possibility that carotenogenic genes have undergone duplication and subfunctionalization through positive selection for the production of specific compounds.

Here, we address the evolution of horizontally transferred carotenoid biosynthetic genes within aphids. We use newly obtained sequences and analyses to address the following questions: At what point in arthropod evolution did the transfer from fungal to animal genome occur, and how many descendant lineages maintain the capability for carotenoid biosynthesis? What group of fungi donated the genes, and does this give clues to the ecological association that may have facilitated the gene transfer? What is the extent of duplication of these genes in different aphid lineages, and does the extent of duplication relate to the profile of carotenoids present? Do duplication events coincide with positive selection in descendant copies, as might be expected if paralogs are being selected for distinct activities? We address these questions by analyzing carotene desaturase sequences newly isolated from a set of species representing major aphid groups and derived from publicly available fungal genome sequences.

Materials and Methods

Study Species

We surveyed 38 insect species including 34 aphid species (Aphididae) plus four species from related insect groups (Adelgidae, Psyllidae, Aleyrodidae) (table 1). The aphids included species from subfamilies and tribes corresponding to several deeply branching lineages, based on von Dohlen and Moran (2000). Among these were representatives of several aphid species that are nearly white in life (*Aspidophorodon longicaudus*, *Hyperomyzus pallidus*, *Macrosiphum diervillae*, *Eucallipterus tiliae*, *Monelliopsis caryae*). For all species, DNA was extracted and used as template in efforts to amplify a region of the gene encoding carotene desaturase. DNA from lab cultures of several species was also used for Southern blot experiments aimed at detecting copies of carotene desaturase. Finally, tissues from some species were assayed for their carotenoid profiles.

Obtaining Sequences of Genes for Carotene Desaturase

Genomic DNA was isolated from fresh or frozen aphids using Qiagen Blood and Tissue kits. The resulting genomic DNA templates were polymerase chain reaction (PCR) amplified with primers complementary to a region of the *A. pisum* carotene desaturase genes. The chosen amplicon is approximately 1,400 bp in length and spans one 71-bp intron in *A. pisum*. The primary pair of primers (TGGAGTTGGTGGTACAGCAG and AGATAATCCTAGTATAGAMCCTTTCCA) corresponds to regions that were highly conserved between all three full-length copies of *A. pisum* carotene desaturase genes, under the assumption that these regions would be most likely to be conserved in other species. In cases in which the initial primer pair failed to produce an amplicon, alternative primers were used (table 2) to maximize retrieval of any carotene desaturase genes present.

To recover sequences from multiple loci that may have amplified with these primers, products of successful PCRs were cloned into Promega pGEM-T Easy vectors, and in average, 20 transformant colonies from each species were picked and their inserts amplified by colony PCR using T7 and SP6 primers. These products were then Sanger sequenced on an ABI3700 sequencer using services at the University of Arizona or at Yale University. Resulting reads for each colony were assembled into a single sequence using Sequencher and manually curated to remove obvious base-calling errors. Subsequently, all sequences with identities $\geq$ 99% were assembled into consensus sequences. In all cases, this process resulted only in the collapse of sequences from the same species. Divergence less than 1% may reflect sequencing error, cloning artifacts, allelic variation, or a combination. In the *A. pisum* genome, carotenoid biosynthetic genes show pairwise divergence of alleles of about 0.13%, whereas paralogous copies are >10% divergent at the nucleotide sequence level (Moran and Jarvik 2010). Thus, although we cannot definitively discriminate allelic variation from divergence between duplicate loci in the newly determined sequences, our cutoff would correctly assign these in *A. pisum*. The number of sequenced PCR products, the number of resulting consensual sequences for each species, and the accession numbers are shown in table 1 and table S1 (Supplementary Material online).

Fungal Sequence

We used available fungal sequences of carotene desaturase in phylogenetic analyses aimed at improving resolution of the source of the transferred genes. Fungal sequences were obtained using blastp with *A. pisum* sequences as queries to retrieve all fungal homologs of carotene desaturase from GenBank, the Joint Genome Institute Fungal Portal (Mycocosm at http://genome.jgi-psf.org/programs/fungi/index.jsf), and the Fungal Genome Initiative at the Broad Institute (http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genome-initiative). Most of fungal sequences obtained were from recently sequenced fungal genomes. We retrieved a total of 52 desaturase sequences derived from 52 fungal species. None of the fungal genome sequences contained multiple loci encoding carotene desaturase.

Analyses to Place the Aphid Sequences among Available Sequences from Fungal Species

Fungal protein sequences were combined with a subset of ten translated aphid desaturase sequences. These ten taxa were selected to represent the full diversity of the aphid sequences, based on preliminary analyses of the aphid

Table 1. Insect Species for Which Carotenoid Biosynthetic Genes Were Studied.

Taxonomic Group	Species Name	PCR[a]	Carot. Assay[d]	South Blot[e]
Aphidinae: Aphidini	*Aphis craccivora*	26/1	+	+
	Aphis nerii	16/2	+	+
	Rhopalosiphum padi	12/5		
	Schizaphis graminum	36/4	+	
Aphidinae: Macrosiphini	*Acyrthosiphon pisum*	4[c]	[c]	+
	Aspidophorodon longicaudus	6/4		
	Brevicoryne brassicae	6/3		
	Diuraphis noxia	6/2	+	
	Hyperomyzus pallidus	4/1		
	Macrosiphum diervillae	5/4		
	Macrosiphum euphorbiae	1/1	+	
	Macrosiphum gaurae	19/7	+	
	Myzus persicae	24/6		+
	Sitobion avenae	15/7	+	+
	Uroleucon ambrosiae	7/2		
	Wahlgreniella nervata red	19/7	+	
	Wahlgreniella nervata green	35/7	+	
Calaphidinae: Calaphidini	*Calaphis* sp.	2/1[b]		
Calaphidinae: Panaphidini	*Myzocallis agrifolicola*	5/3	+	
	Eucallipterus tiliae	5/1	+	
	Monelliopsis caryae	5/1		
Drepanosiphinae	*Drepanaphis utahensis*	1/1	+	
Eriosomatinae: Eriosomatini	*Eriosoma lanigerum*	9/2	+	
Eriosomatinae: Fordini	*Geopemphigus floccosus*	7/3	+	
	Melaphis rhois	8/1		
	Schlechtendalia chinensis	9/1		
Eriosomatinae: Pemphigini	*Pemphigus populiramulorum*	4/1		
Eriosomatinae: Prociphilini	*Prociphilus* sp.	4/1		
Hormaphidinae: Hormaphidini	*Hamamelistes spinosus*	5/3		
	Hormaphis hamamelidis	5/3	+	
Chaitophorinae: Chaitophorini	*Chaitophorus populifolii*	8/2		
	Chaitophorus stevensis	8/3		
Lachninae: Cinarini	*Cinara* sp.	3/2	+	
	Cinara pinea group	4/1		
	Cinara ponderosae	4/2		
	Tuberolachnus salignus	1/1[b]		
Adelgidae	*Adelges cooleyi*	1/1[b]	+	
	Adelges laricis	1/1[b]		
Aleyrodidae	*Bemisia tabaci*	0/0	+	
Psyllidae	*Pachypsylla venusta*	0/0	+	

[a] No. of sequences/No. of copies; for accession numbers and collection information, see supplementary table S2 (Supplementary Material online).
[b] PCR product amplified with primers PD-D3F/PD-D7R.
[c] Moran and Jarvik (2010).
[d] Carotenoid profile assayed, see supplementary figure S2 (Supplementary Material online).
[e] Southern blot analysis, see supplementary figure S1 (Supplementary Material online).

sequences. A data set consisting of 63 taxa, including one desaturase sequence of bacterial origin as an outgroup, was aligned at a protein level in server-based program MAFFT (Multiple Alignment using Fast Fourier Transformation, http://mafft.cbrc.jp/alignment/server/index.html), using the E-INS-i algorithm with default parameters. The raw alignment was manually corrected in program BioEdit (Hall 1999) and further processed in GBlocks application (Castresana 2000) in order to remove unreliably aligned regions containing gap positions. The resulting alignment was analyzed using maximum likelihood (ML) and Bayesian inference (BI). ML-based analyses and 100 nonparametric bootstrap replicates were performed in the PhyML program (Guindon and Gascuel 2003) with the best fitting model LG + Γ selected in ProtTest 3 (Darriba et al. 2011) and parameters estimated from the data. Bayesian analysis was performed with the Whelan

Table 2. List of Oligonucleotides Used for PCR and Southern Hybridization.[a]

Name	Forward Sequence (5′–3′)	Name	Reverse Sequence (5′–3′)	Amplicon Length (bp)
torF19584	TGGAGTTGGTGGTACAGCAG	torR20949	AGATAATCCTAGTATAGAMCCTTTCCA	1365
torF19584	TGGAGTTGGTGGTACAGCAG	torR20693	CGATGYGRCTRGGWACGT	1109
torF20391	GAYGACAAMGGWGTGGCGA	torR20684	CGRCTRGGWACGTTMACRTAAA	293
PD-D3F	CCNAGDATNGANCCYYTCCA	PD-D7R	GCNGARGGNATHTGGTAYCC	689
Probe100F	TTYGATCAAGGHCCATCATT	Probe601R	CCTCCTTTYGGRTACCADAT	499

[a] Amplicon length corresponds to the gene region for carotene desaturase from *Acyrthosiphon pisum* genomic contigs.

and Goldman model of protein evolution implemented in MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003) and following parameter settings: rates = gamma, ngen = 4,000,000, samplefreq = 100, and printfreq = 100. Other Markov chain Monte Carlo setting and prior distributions were set at the default values.

Analyses to Detect Recombination of the Gene Copies

The intron sequence was manually spliced out, and the set of 102 sequences was aligned as described above. Potentially pseudogenized sequences containing frameshifts or premature stop codons were temporarily removed from the matrix and a reduced alignment was tested for the presence of recombination breakpoints using Single Breakpoint Recombination and Genetic Algorithms for Recombination Detection (Kosakovsky Pond et al. 2006) algorithms as implemented in the program DataMonkey (Delport et al. 2010).

Tests of Selection for the Clade of Carotene Desaturases within Aphids

As the first nonrecombinant region (627 bp) spanned only 36 bp of several shorter sequences, including outgroup sequences of *Adelges laricis* and *Adelges cooleyi*, only the second matrix consisting of 413 nt position was further analyzed. Due to the short length of this matrix, data were analyzed at a nucleotide level only. Analyses were performed as for the initial phylogeny described above with the exception of a few steps concerning the use of nucleotide characters. General time reversible (GTR) + Γ was determined as the best fitting model of molecular evolution in program jModelTest (Posada 2008) and used in ML-based analyses. Bayesian probability was inferred using the same model implemented in MrBayes version 3.1.2. (Ronquist and Huelsenbeck 2003) with the following parameter settings: nst = 6, rates = gamma, ngen = 20,000,000, samplefreq = 100, and printfreq = 100.

To test for positive selection, we analyzed a set of intact partial sequences from the second nonrecombinant matrix region in order to detect changes in selective forces over time. Programs CodeML from the PAML software package (Yang 2007) and DataMonkey (Delport et al. 2010) were used to estimate omega, the ratio of nonsynonymous changes per nonsynonymous site to synonymous changes per synonymous site, or d*N*/d*S*, for particular nodes on phylogenetic trees for the gene family. In CodeML, the unconstrained free-ratio model allowing independent omega values for every branch was used in attempts to estimate variation in omega across the phylogeny. To further investigate when and how selection pressure varied over the evolutionary history, we applied the GA-Branch (Genetic Algorithm-Branch) method (Kosakovsky Pond and Frost 2005a) implemented in the program DataMonkey (Delport et al. 2010). Because the full data set with 80 sequences exceeded capabilities of the program, we used a reduced set of 8 taxa and 29 copies in our tests. To exclude the possibility that choice of copies or taxa was biasing results, we performed several analyses with different sets of sequences to determine if the same branches consistently gave similar estimates of omega. We generated these data sets with the aim of attaining an optimum sequence divergence as suggested by Yang (2002).

Residues likely to change the substrate specificity or site of desaturation are not defined for carotene desaturase. The NADH-binding motif is known (Pecker et al. 1992) and is a nonrecombinant region in our data set. As a second test for potentially detecting positive selection, the region containing the NADH-binding motif was analyzed using site-specific tests SLAC and FEL (Kosakovsky Pond and Frost 2005b) implemented in DataMonkey (Delport et al. 2010).

Southern Blots

Genomic DNA of five aphid species (table 1) was digested overnight with HincIII (New England Biolabs), separated in a 1% agarose gel, and transferred onto a nylon membrane (Roche Diagnostics) using rapid alkaline transfer protocol (Reed and Mann 1985). Preparation of labeled probes targeting a conserved gene region (table 2), Southern hybridization, and detection of hybridization signals using a Image Quant LAS-4000 system (GE Healthcare) were carried out with a DIG DNA labeling and detection kit (Roche Diagnostics).

Carotenoid Profiles

Whole bodies of a subset of species were analyzed for carotenoid profiles using high-performance liquid chromatography at Craft Technologies, as previously described (Moran and Jarvik 2010).

Results

Retrieval of Carotene Desaturase Gene Sequences from Insect Samples

All 34 species of Aphididae yielded PCR products. Following cloning, each species yielded from one to seven distinct DNA sequences encoding carotene desaturase. A total of 98 distinct sequences were obtained, based on collapsing sequences with <1% divergence and starting with a set of 336 sequences. In most cases, PCR product, corresponding to about 1,100 nt and encoding 330 amino acids, was retrieved. A few species, including the adelgids, gave products only for alternative primers that amplified a shorter region and that did not span an intron in the *A. pisum* sequences. This could reflect sequence divergence at primer sites or intron expansion, as seen in *E. tiliae* and *M. caryae*, in which intron lengths exceed 600 bp. Sequences obtained using all primer pairs showed unambiguous homology to carotene desaturase and were most similar to those from *A. pisum*, based on blastx searches of existing protein databases. Thus, aphids from diverse lineages corresponding to different subfamilies and tribes possess carotene desaturase.

Adelgids and phylloxerids are the closest relatives of Aphididae (von Dohlen and Moran 1995, 2000). Our two adelgid samples yielded a product only for the shorter PCR amplicon of about 700 bp, and the sequence showed

clear homology to carotene desaturase. We could not obtain amplicons from samples from Aleyrodidae or Psyllidae, groups more distantly related to aphids within the suborder Sternorrhyncha. Failure to obtain a PCR product is not strong evidence of absence of these genes, so presence of carotene desaturase genes in other Sternorrhyncha, cannot be excluded. However, blast searches of 19,598 expressed sequence tags (ESTs) available in GenBank (12 May 2011) for another psyllid species, *Diaphorina citri*, did not yield any significant hits for carotene desaturase genes. Based on frequency of carotene desaturase transcripts in the *A. pisum* EST set (118 of 108,686 ESTs), we would expect about 20 *D. citri* ESTs if the genes were expressed at the same level, suggesting that they are absent from this psyllid species.

Each aphid species yielded from one to seven distinct carotene desaturase genes, under our criterion of >1% divergence for distinct copies. Although we sequenced multiple clones in an effort to sample the diversity of copies from each species, it is likely that some copies did not amplify, did not clone, or were not selected for sequencing. Thus, our data may underestimate the diversity of copies present within some aphid species. However, in some cases, we invested considerable effort into obtaining sequences for all paralogs, through the use of alternative primers and through extensive sequencing of cloned copies. For example, for *Aphis craccivora*, 26 clones were sequenced. These were nearly identical and collapsed into a single sequence under our criterion of >99% identity. In some cases, such as *Aphis nerii*, *A. longicaudus*, and *Macrosiphum gaurae*, two very similar copies (with identity only slightly <99%) were retrieved; possibly these represent alleles.

Assessing Number of Carotene Desaturase Genes

To further determine whether different aphid species contain different numbers of carotene desaturase genes, we performed Southern hybridizations with five species that differed in number of copies detected by PCR (supplementary fig. S1, Supplementary Material online). We obtained evidence for four copies in *A. pisum* clone LSR1, corresponding to the known number of copies in the completely sequenced genome (Moran and Jarvik 2010). For *A. craccivora* and *A. nerii*, one and two copies were found, corresponding to the results from cloning and sequencing. We obtained four to five bands for both *Myzus persicae* and *Sitobion avenae*, for which PCR and sequencing yielded six distinct sequences. Thus, differences in copy number among species are supported in cases in which more extensive sequencing of clones was carried out. Furthermore, the results suggest a pronounced difference in the extent of duplication for Aphidini and Macrosiphini species, including *A. pisum*, *M. gaurae*, *Wahlgreniella nervata*, and *M. persicae*, which contain many copies.

Certain aphids are very pale in life, suggesting the absence of carotenoids; however, some of the more saturated carotenoids, such as phytoene or zeta-carotene, have almost no color (Britton et al. 2004). Even very pale aphid species contained at least one apparently intact copy of carotene desaturase, based on our amplification and sequencing of most of the coding region.

Detection of Pseudogenes for Carotene Desaturase

Pseudogenes were detected in numerous species, based on the presence of obvious deletions or base changes that interrupted the reading frame and implied an inactivated gene. Detected pseudogenes were always very closely related to sequences that appeared to be from intact genes, indicating recent inactivation events. For example, all nine pseudogene sequences of *A. nerii* collapsed to *A. nerii* copy A according to our <1% divergence criterion, and all contained the same base substitution resulting in an early stop codon. Because we sequenced only part of the gene and did not measure expression, we might not have detected all cases of gene inactivation.

Phylogeny of Carotene Desaturases from Aphids and Fungi

Following preliminary analyses of the aphid-derived desaturase sequences, a set of ten aphid sequences representing the diversity of the larger set was used in an analysis that included all available fungal sequences for carotene desaturase (fig. 1). In topologies inferred by both ML (fig. 1) and BI (supplementary fig. S2, Supplementary Material online), the aphid sequences form a strongly supported clade, which branches outside of the Basidiomycota or Ascomycota. Among available sequences, the aphid sequences are closest to those from the Mucoromycotina. However, many basal fungal lineages are not represented among sequenced genomes, and the branch leading to the aphid clade is long. Thus, the closest relatives of the aphid sequences may fall within a fungal group not represented in our tree, possibly within the Entomophthorales which contains species parasitic in arthropods. The adelgid sequence belonged to the aphid clade with strong support in all analyses but was short and thus not used in most analyses.

These observations are most consistent with the hypothesis of a single acquisition of this gene in an ancestor of extant Aphidoidea and Adelgidae from a fungus outside of the Basidiomycota or Ascomycota.

Ongoing Duplications of Carotene Desaturase Genes within Aphids

For the full set of sequences from aphids and adelgids, analyses based on ML and BI (Materials and Methods) gave nearly identical results. However, even after 20,000,000 generations, the BI-derived topology was not fully resolved due to the limited amount of information retained. Thus, we present only the ML phylogeny (fig. 2). The only substantial difference in topology in the BI tree involves the position of two *Chaitophorus* species, which cluster at the base of Eriosomatinae taxa rather than at the base of the whole Aphididae as in the ML tree. This difference reflects the

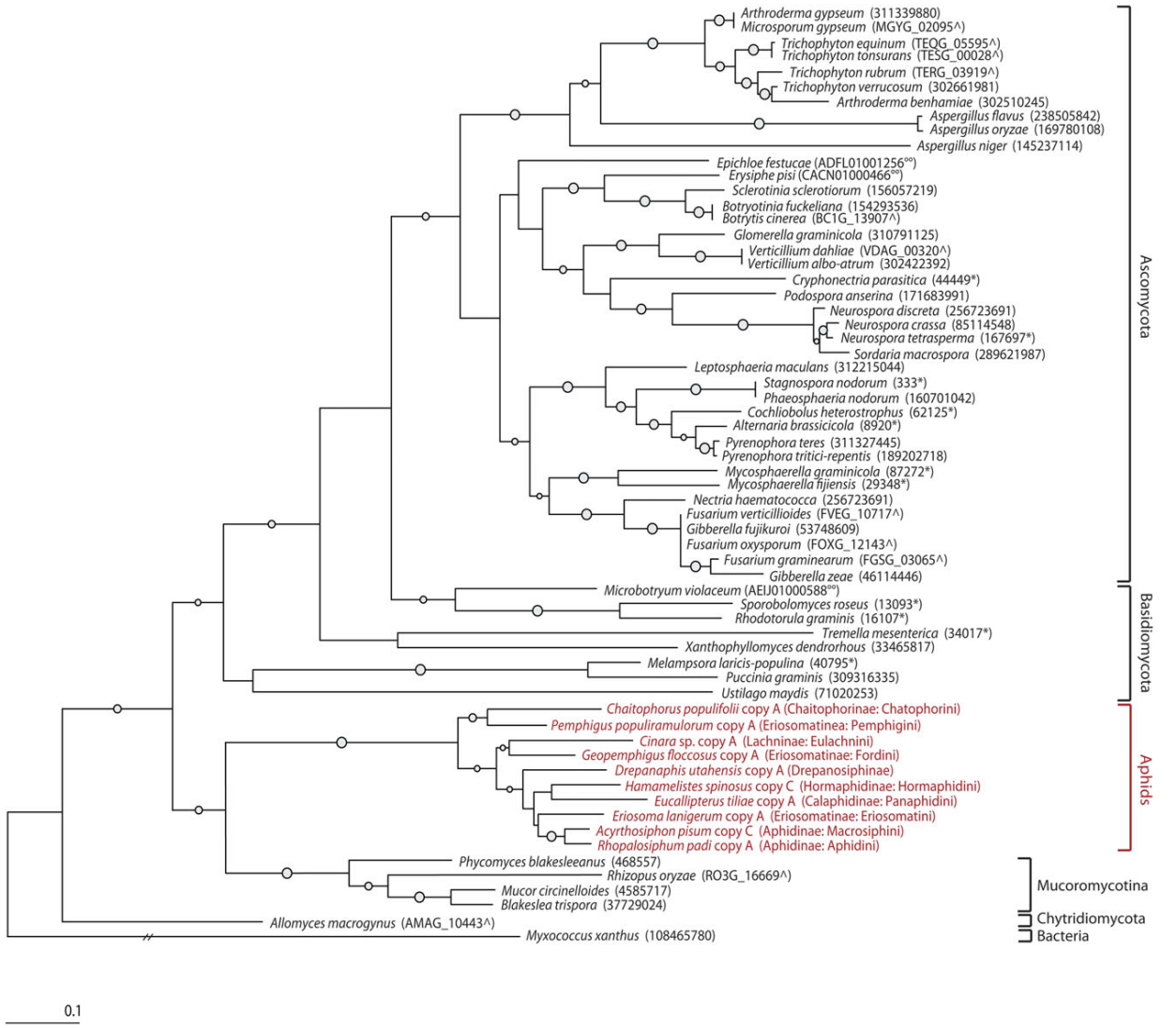


Fig. 1. ML phylogenetic tree for carotene desaturase proteins from fungi and from representative aphids. Accession numbers other than GenBank GI are designated as follows: GenBank WGS (open circles), the Fungal Genome Initiative at the Broad Institute protein ID (^), the Joint Genome Institute Fungal Portal locus ID (*). Solid circles indicate bootstrap values above 50 and are scaled accordingly. Accession numbers for aphid taxa from top to bottom: JN022723, JN022746, JN022711, JN022731, JN022727, JN022738, JN022781, JN022728, XM_001946654, and JN022748.

occurrence of a long branch for these sequences and does not affect general conclusions for evolution of the carotene desaturase genes. As noted, the tree may not contain all copies present in all species included in the analysis, due to failure of some copies to amplify or to be sequenced. Nonetheless, several conclusions are possible.

First, gene duplications date to a deep node in the aphid desaturase tree, implying that these genes have undergone duplication since the origin of Aphidinae or earlier. For example, *A. pisum* copy A corresponds to a cluster within the Macrosiphini that diverges from other Macrosiphini copies at a deep node. However, the most basal node in the tree separates aphids belonging to different subfamilies and thus does not appear to correspond to a duplication event. We note that the rooting is based on the shorter sequences obtained from the adelgid samples and is thus not certain. Fungal sequences were too divergent to be used for rooting.

Second, duplications are ongoing in many lineages, and many species contain closely related sister sequences present within the same sample. Although some of these could be allelic differences, the divergence is always >1% (since sequences closer than this were collapsed) and sometimes much more. Allelic differences are expected to be less than 1%, at least based on data from *A. pisum* (Moran and Jarvik 2010). Other evidence that at least some close copies represent duplicated loci rather than alleles comes from the retrieval of more than two close variants from samples originating from a single aphid clone grown in the lab. For example, *Brevicoryne brassicae*, *M. persicae*, *Rhopalosiphum padi*, and *Schizaphis graminum* were all reared as diploid asexual clones derived from a single female and thus

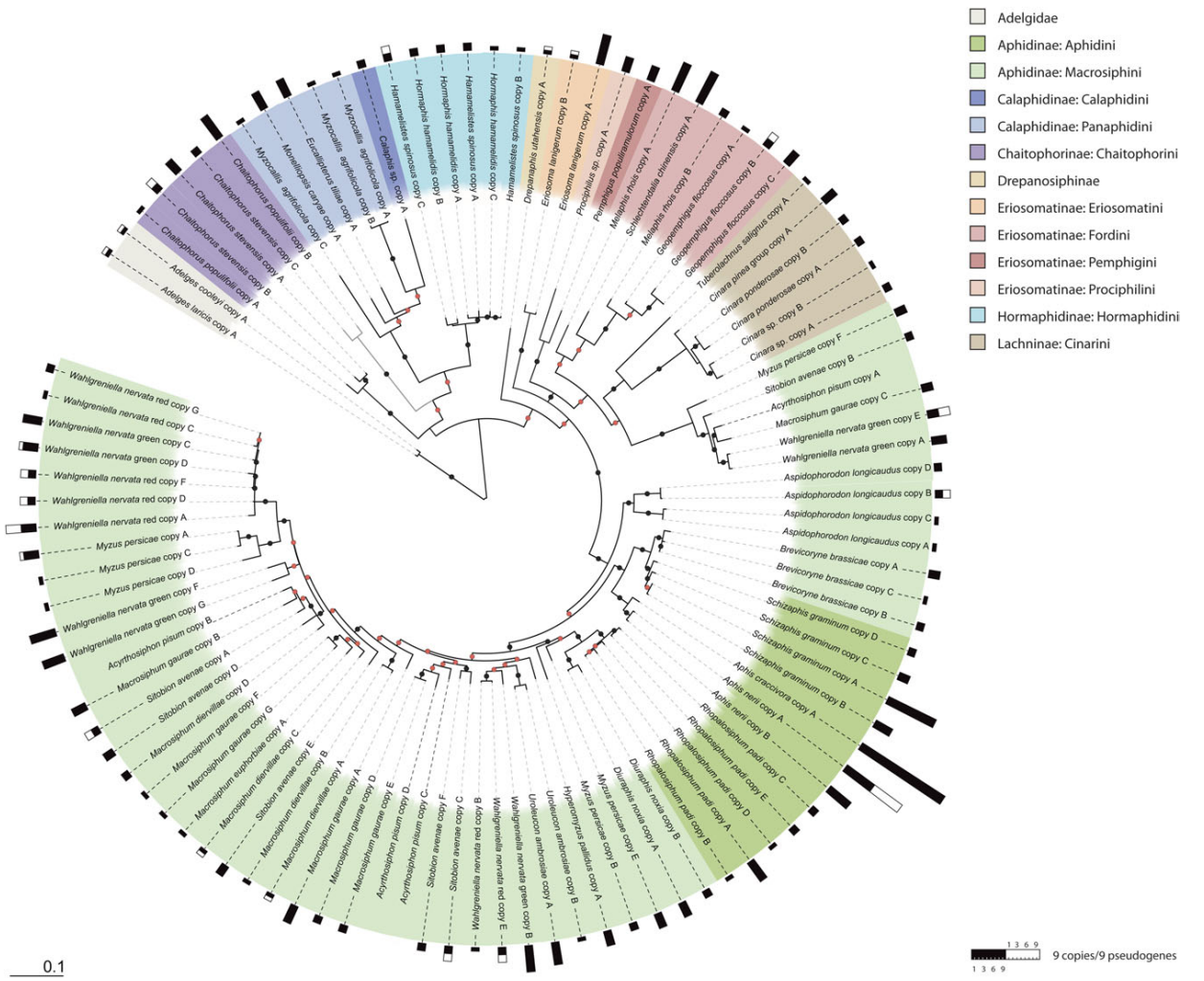


Fig. 2. ML phylogenetic tree for carotene desaturase genes from aphids. Black bars represent the total number of collapsed sequences according to <1% divergence rule. White bars stands for the proportion of sequences that are pseudogenized. Solid circles indicate bootstrap values above 50 and are scaled accordingly. The ambiguous position of *Chaitophorus* sequences is depicted in gray color. Dark dashed lines highlight the position of the four carotene desaturase copies from the *Acyrthosiphon pisum* genome.

can possess at most two alleles per locus; yet each contains three to five close copies. In some lineages, several unique duplications have occurred; examples include *A. longicaudus*, *B. brassicae*, *M. persicae*, *R. padi*, and *S. graminum*. And, in certain deeper clades of aphids, unique duplications have given rise to radiations confined to particular aphid groups. This can be observed most dramatically in the Macrosiphini.

Thus, duplication of genes for carotene desaturase has been ongoing in aphids. Many detected duplicates are closely related, but others are ancient. Persistence of ancient duplicates is more evident in some lineages than in others, based on the evidence for only one or few loci in certain species such as *A. craccivora* and *A. nerii*. Together, these observations suggest a high rate of duplication and a high rate of inactivation of recently duplicated copies. At the same time, certain lineages show retention of more ancient duplicates, as in the case of the divergent copies present within genomes of several species of Macrosiphini (*A. pisum*, *M. persicae*, *M. gaurae*, *S. avenae*, and *W. nervata*).

Occurrence of Recombination among Carotene Desaturase Families

According to the position of the identified breakpoint, the initial alignment of all 102 sequences was split into two matrices, of 627 and 413 bp in length, respectively. The first matrix contained fewer taxa since not all species were successfully amplified for this region. The 82 nonpseudogenized sequences from the second matrix were further tested for positive selection in alternate reduced data sets.

Assessing Positive and Purifying Selection on Aphid Carotene Desaturase Genes

We performed several analyses designed to detect variation in the average selective forces, measured as omega, over the branches on phylogenies for distinct reduced sets of

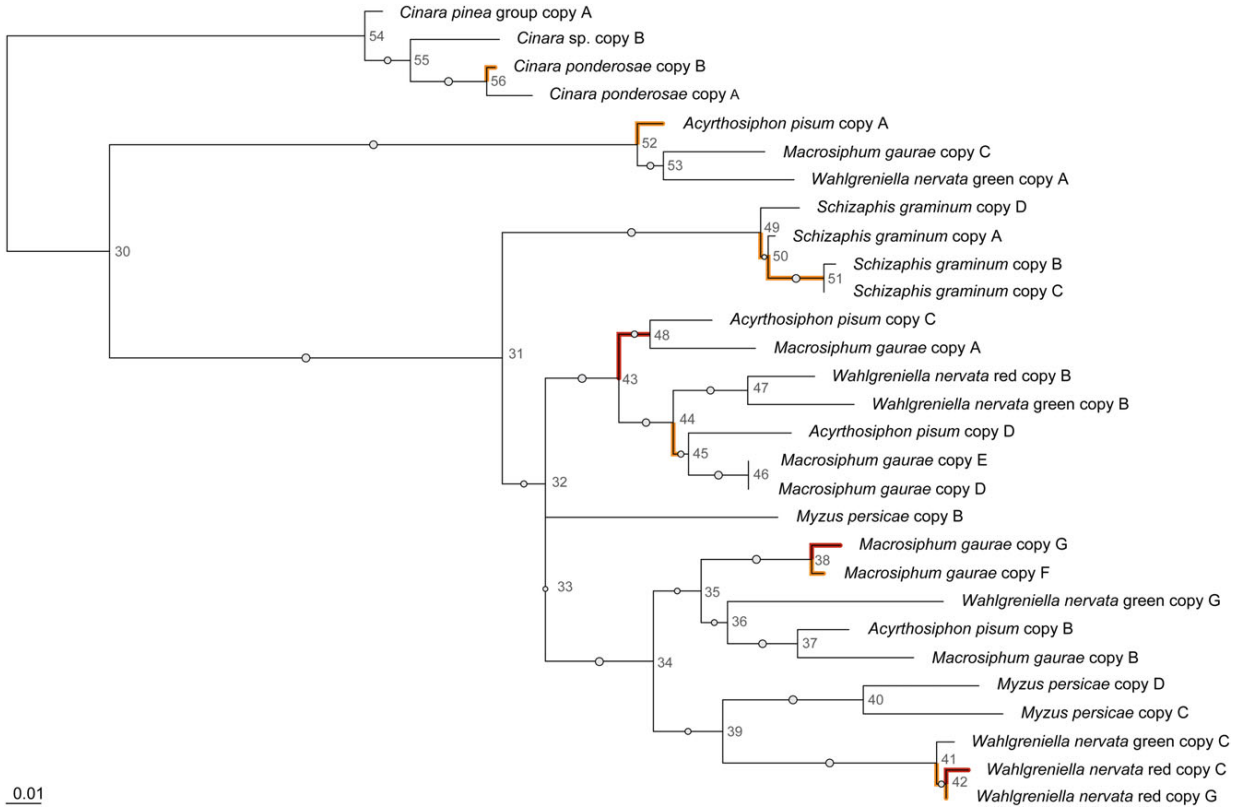


Fig. 3. ML-based topology for aphid carotene desaturases highlighting branches with omega (dN/dS) > 1, indicating positive selection. Red color indicates branches selected in several independent GA-Branch analyses as well as CodeML analysis. Orange highlights branches with omega >1, calculated with the free-ratio model in CodeML (Materials and Methods). Solid circles indicate bootstrap values above 50 and are scaled accordingly. Internal nodes are numbered.

carotene desaturase genes. All of our analyses supported variation in selection among branches. Several independent GA-Branch analyses assigned four to five omega categories to branches of our phylogenies (Results not shown). Omega values above were consistently retrieved for three percent of the branches, indicating that sporadic positive selection has acted on certain copies of this gene. Figure 3 presents the consensus of results from CodeML and GA-Branch analyses, where three nodes are followed by branches having omega estimates statistically greater than 1; all of these follow duplication events. Positive selection is supported for both deep and recent nodes. For example, the branch leading to node 48, a clade consisting of sequences from Macrosiphini (*A. pisum* copy C + *M. gaurae* copy A), shows a strong signature of positive selection. Examples of recent duplications followed by elevated omega values include several cases in which duplicates are confined to one species within our sample (*Cinara ponderosae*, *S. graminum*, *M. gaurae*, and *W. nervata*).

Site-specific tests performed on the nonrecombinant region containing the NADH-binding motif did not yield statistically significant evidence of positive selection. Because our tests were limited to a somewhat short alignment, and because our sampling of taxa and gene copies is incomplete, these tests should be regarded as preliminary, but they strongly suggest repeated subfunctionalization associated with adaptive evolution during the evolution of this gene family.

Carotenoid Composition of Aphid Species

Few aphid species have been assayed for carotenoid contents, raising the question of whether these genes continue to function in carotenoid production in diverse aphid lineages and whether the presence of multiple copies correlates with production of diverse carotenoid compounds. To address this, we obtained carotenoid profiles for a subset of species for which sufficient material could be obtained (fig. 4). All species of Aphididae, including very pale aphids, contain carotenoids of the C40 type. This observation strongly supports the view that, in aphids, carotene desaturase as well as phytoene synthase/carotene cyclase continue to function in the biosynthesis of carotenoids of the same types observed in fungi. In pale species, such as *E. tiliae* and *Pemphigus betae*, carotenoid profiles are dominated by relatively colorless C40 carotenoids.

The adelgid sample also contained C40 carotenoids, as did the whitefly and psyllid samples. Xanthins, which require a carotene hydroxylase step and which are expected as dominant carotenoids in most plant tissues, most likely represent carotenoids ingested with food. They are largely absent from species of Aphididae (fig. 4), supporting the hypothesis that aphids produce their own carotenoids,

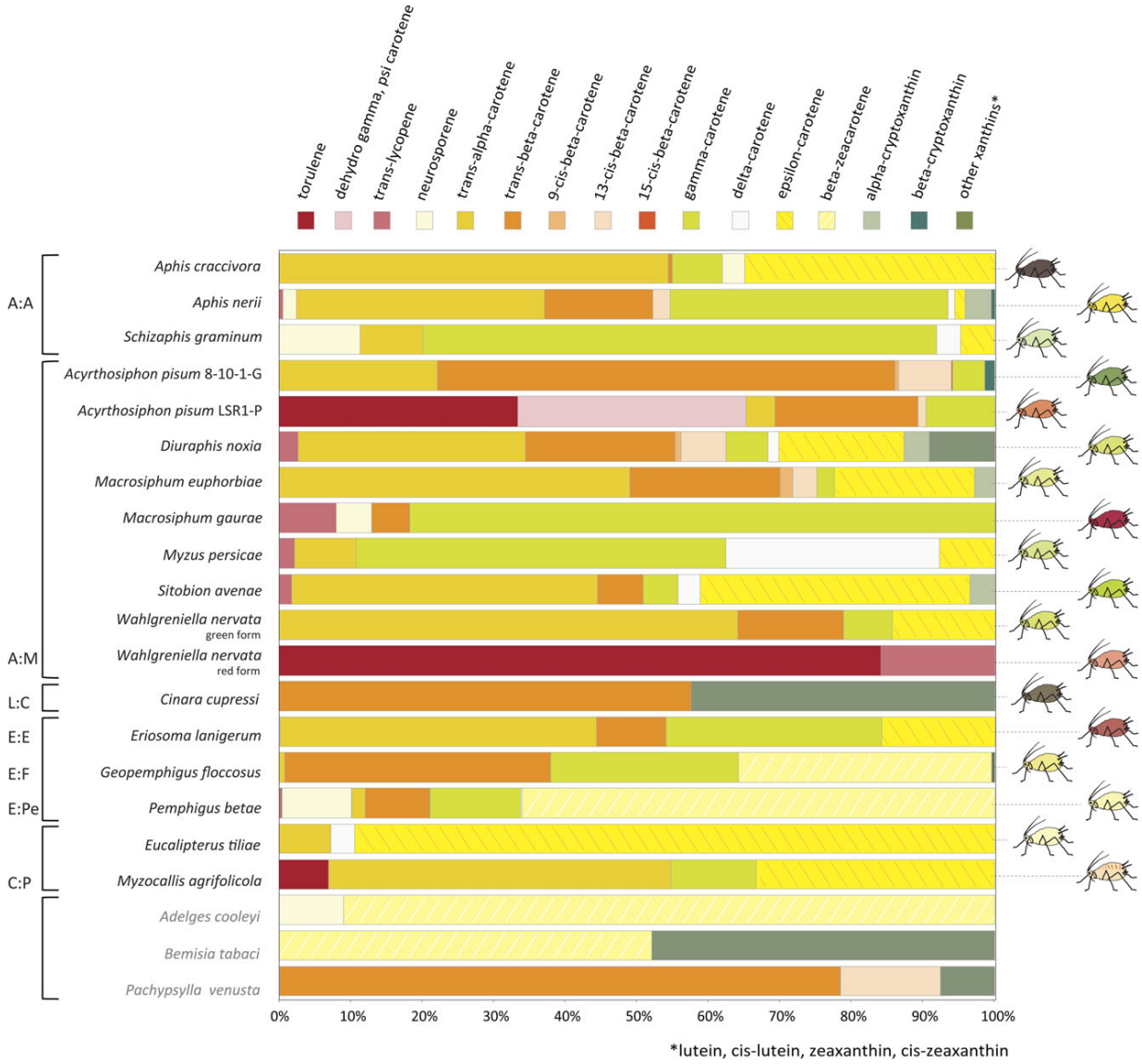


Fig. 4. Profiles of carotenoids obtained for different aphid species. Width of the bar indicates proportional representation among carotenoids detected in samples. Approximate color in life for each species is presented on the right. Xanthins, which are of likely plant origin, are presented as green colors on the right. Brackets and letters along the left side indicate the higher taxonomic groupings of the hosts and correspond to the taxonomic information presented in table 1.

using their own enzymatic machinery. Small amounts of xanthins are found in *Diuraphis noxia*, a species that induces degradation of host plant cells, possibly releasing carotenoid into the ingested plant sap, in *Cinara cupressi*, and in the whitefly and psyllid samples.

In the case of one copy of carotene desaturase in *A. pisum*, a distinct role in the production of torulene has been demonstrated (Moran and Jarvik 2010). For other cases, specific compounds cannot be linked to particular gene copies, but the overall picture is that aphids produce a variety of C40 compounds and that C40 compounds dominate among carotenoids present in aphid tissues. Red body color in different aphid species can reflect either the presence of torulene (*A. pisum*, *W. nervata*), lycopene (*M. gaurae*), or both of these red compounds. The presence of C40 carotenoids in psyllids and whiteflies is unexplained and raises the possibility that undetected carotenogenic genes are present in these species. We failed to recover carotene desaturase sequences with degenerate PCR of psyllid and whitefly DNA samples, but these efforts do not provide definitive evidence that such sequences are absent.

Discussion

The presence of multiple loci for carotenoid biosynthesis is highly unusual within sequenced genomes of plants, fungi, and prokaryotes, yet most species of aphids appear to have multiple copies of carotene desaturase genes. These

findings on carotene desaturase possibly are illustrative of the evolutionary processes dominating throughout the aphid genome: the *A. pisum* genome sequence revealed that aphid evolution has involved an exceptionally high level of gene family expansion through duplication (International Aphid Genomics Consortium 2010; Ollivier et al. 2010). For the genome overall, divergences of paralogs are as large as those of orthologs from distantly related aphid species, indicating that duplication resulting in expansion of gene families has been ongoing since near the time of origin of extant aphids or even longer (International Aphid Genomics Consortium 2010). Our results reveal a similar history for carotene desaturase, as indicated by the occurrence of duplication events deep in the clade of aphid gene copies (fig. 2). A high rate of duplications in many lineages is consistent with the finding of many pseudogene copies and suggests that duplication is often soon followed by nonfunctionalization.

However, some paralogs are retained within genomes for long periods; for example, the basal divergence of the paralogs present in *A. pisum* would predate the divergence of Eriosomatinae, Drepanosiphinae, Lachninae, and Aphidinae, which form basal lineages among extant aphids and which are estimated to have diverged at least 150–80 Ma, based on fossil and molecular evidence (von Dohlen and Moran 2000). Furthermore, paralogs retained in genomes often undergo positive selection for amino acid replacements (fig. 3). Thus, it appears that different copies have specialized to particular functions. Since expression confined to different developmental stages or tissues does not appear likely, functional differences between copies are more likely to involve different substrate specificity and production of specific carotenoid types. In the case of *A. pisum*, one desaturase copy was shown to be necessary for production of torulene, the basis of red body color (Moran and Jarvik 2010). We note that the other red/green color polymorphic species included in our samples, *M. persicae*, *W. nervata*, *S. avenae*, and *M. gaurae*, also have large number of desaturase copies. This suggests that the multiple copies are linked to the capacity for evolving novel carotenoid profiles, potentially linked to particular ecological circumstances. Carotene desaturase of bacteria has been used in experimental evolution studies, which have shown that minor changes in the enzyme can lead to the production of novel carotenoids including torulene (Schmidt-Dannert et al. 2000).

The relationship between the phylogenies for the desaturase gene family and for aphid lineages is complex, and our sampling of aphid species is limited. The cluster containing desaturase copies from species in Eriosomatinae, Lachninae, and Aphidinae: Macrosiphini (fig. 2) spans diverse aphid lineages that diverged over a relatively short period estimated to be at least 100 Ma (von Dohlen and Moran 2000; von Dohlen et al. 2006). We note that the Eriosomatinae does not form a clade in our trees and that this is similar to results in other molecular phylogenetic analyses based on single copy aphid genes (e.g., Martinez-Torres et al. 2001; Ortiz-Rivas et al. 2004; Zhang and Qiao 2008).

Our analyses provide a picture of the evolution of carotene desaturase genes in aphids following their acquisition from a fungus. They reveal that these genes have a single origin in a shared ancestor of aphids and adelgids, are ubiquitous in all living aphids, and have diversified through repeated bouts of duplication and selection. This suggests an important role of carotenoids in aphid biology and diversification.

Supplementary Material

Supplementary figures S1 and S2 and Supplementary table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Tyler Jarvik for laboratory work on initial stages of this project, Carol von Dohlen, Robert Foottit, and Eric Maw for providing aphid samples and identifications, Kim Hammond for her help with live aphid colonies and sampling, and Neal Craft and Craft Technologies for the carotenoid assays. This work was supported by the United States National Science Foundation (DEB-0723472, DEB-1106195 to N.A.M.). E.N. was supported on a Fulbright fellowship and by the Grant Agency of the University of South Bohemia (135/2010/P (a)) and the Grant Agency of the Czech Republic (206/09/H026).

References

Britton G, Liaaen-Jensen S, Pfander H, editors. 2004. Carotenoids handbook. Basel (Switzerland): Birkhäuser-Verlag.

Britton G, Liaaen-Jensen S, Pfander H, editors. 2006. Carotenoids: volume 4: natural functions. Basel (Switzerland): Birkhäuser-Verlag.

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

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.

Delport W, Poon AF, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.

Gallagher CE, Matthews PD, Li F, Wurtzel ET. 2004. Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses. *Plant Physiol*. 135:1776–1783.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 41:95–98.

International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol*. 8:e1000313.

Klassen JL. 2010. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *PLoS One*. 5:311257.

Kosakovsky Pond SL, Frost SDW. 2005a. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol*. 22:478–485.

Kosakovsky Pond SL, Frost SDW. 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 22:1208–1222.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol*. 23:1891–1901.

Martinez-Torres D, Buades C, Latorre A, Moya A. 2001. Molecular systematics of aphids and their primary endosymbionts. *Mol Phylogenet Evol*. 20:437–449.

Moran NA, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328:624–627.

Ollivier M, Legeal F, Rispe C. 2010. Comparative analysis of the *Acyrthosiphon pisum* genome and expressed sequence tag-based gene sets from other aphid species. *Insect Mol Biol*. 19:33–45.

Ortiz-Rivas B, Moya A, Martnez-Torres D. 2004. Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene. *Mol Phylogenet Evol*. 30:24–37.

Pecker I, Chamovitz D, Linder H, Sandmann G, Hirschberg J. 1992. A single polypeptide catalyzing the conversion of phytoene to zeta-carotene is transcriptionally regulated during tomato fruit ripening. *Proc Natl Acad Sci U S A*. 89:4962–4966.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25:1253–1256.

Reed KC, Mann DA. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res*. 25:7207–7221.

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

Schmidt-Dannert C, Umeno D, Arnold FH. 2000. Molecular breeding of carotenoid biosynthetic pathways. *Nat Biotechnol*. 18:750–753.

von Dohlen CD, Moran NA. 1995. Molecular phylogeny of the Homoptera: a paraphyletic taxon. *J Mol Evol*. 71:689–717.

von Dohlen CD, Moran NA. 2000. Molecular data support a rapid radiation of aphid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol J Linn Soc*. 71:689–717.

von Dohlen CD, Rowe CA, Heie OE. 2006. A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. *Mol Phylogenet Evol*. 38:316–329.

Yang Z. 2002. Inference of selection from multiple species alignments. *Curr Opin Genet Dev*. 12:688–694.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Zhang HC, Qiao GX. 2008. Molecular phylogeny of Pemphiginae (Hemiptera: Aphididae) inferred from nuclear gene EF-1*a* sequences. *Bull Entomol Res*. 98:499–507.

Paper 3:

Thairu, M. W., Skidmore, I. H., Bansal, R., **Nováková, E**., Hansen, T. E., Li-Byarlay, H., ... & Hansen, A. K. (2017). Efficacy of RNA interference knockdown using aerosolized short interfering RNAs bound to nanoparticles in three diverse aphid species. *Insect molecular biology*, *26*(3), 356-368.

Insect Molecular Biology

Insect Molecular Biology (2017) **26**(3), 356–368

Efficacy of RNA interference knockdown using aerosolized short interfering RNAs bound to nanoparticles in three diverse aphid species

M. W. Thairu *¶, I. H. Skidmore*, R. Bansal†, E. Nováková*, T. E. Hansen*, H. Li-Byarlay‡, S. A. Wickline§ and A. K. Hansen*

**Department of Entomology, University of Illinois, Urbana-Champaign, IL, USA;* †*Department of Entomology, The Ohio State University, Wooster, OH, USA;* ‡*Department of Entomology, North Carolina State University, Raleigh, NC, USA; and* §*School of Medicine, Washington University in St. Louis, MO, USA*

Abstract

RNA interference (RNAi) has emerged as a promising method for validating gene function; however, its utility in nonmodel insects has proven problematic, with delivery methods being one of the main obstacles. This study investigates a novel method of RNAi delivery in aphids, the aerosolization of short interfering RNA (siRNA)–nanoparticle complexes. By using nanoparticles as a siRNA carrier, the likelihood of cellular uptake is increased, when compared to methods previously used in insects. To determine the efficacy of this RNAi delivery system, siRNAs were aerosolized with and without nanoparticles in three aphid species*: Acyrthosiphon pisum, Aphis glycines* **and *Schizaphis graminum*. **The genes targeted for knockdown were** *carotene dehydrogenase* **(***tor***), which is important for pigmentation in** *Ac. pisum*, **and** *branched chain-amino acid transaminase* **(***bcat***), which is essential in the metabolism of branched-chain amino acids in all three aphid species. Overall, we observed modest gene knockdown of** *tor* **in** *Ac. pisum* **and moderate gene knockdown of** *bcat* **in** *Ap. glycines* **along with its associated phenotype. We also determined that the nanoparticle emulsion significantly increased the efficacy of gene knockdown. Overall, these results suggest that the aerosolized siRNA–nanoparticle delivery method is a promising new high-throughput and non-invasive RNAi delivery method in some aphid species.**

Keywords: RNAi delivery, siRNA, gene knockdown, ***branched-chain amino acid transaminase (bcat)*****.**

First published online 17 March 2017.

Correspondence: Margaret W. Thairu, Department of Entomology, University of Illinois, Urbana-Champaign, IL 61801, USA. E-mail: thairu2@illinois.edu

¶*Present address:* Department of Parasitology, Faculty of Science, University of South Bohemia and Institute of Parasitology, Biology Centre, Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic.

Introduction

Currently, the biggest challenge for the field of genomics is the functional characterization of genes and linking them to organismal phenotypes. Over the last decade, there has been substantial progress in the sequencing of eukaryotic genomes. Presently, nearly 720 animal genomes including numerous insect, mammal and avian species have been fully sequenced and/or are in draft form (NCBI, 2017). Gene prediction and annotation for these eukaryotic genomes has relied primarily on bioinformatics approaches utilizing *ab initio* gene models and homology. However, many of these computational predictions remain to be functionally validated. In nonmodel organisms, there is the additional challenge of characterizing lineage-specific genes. Consequently, there is a need for functional genomic techniques that can be easily used in both model and nonmodel systems.

One promising approach for validating gene function in animals is RNA interference (RNAi), a type of post-transcriptional gene silencing that targets a specific mRNA. By introducing double-stranded RNA (dsRNA) or short interfering RNA (siRNA) the RNAi pathway can be induced to silence a gene of interest (Fig. S1). The RNAi pathway is highly conserved across eukaryotes and within insects, including aphids (Shabalina & Koonin, 2008; Bellés, 2010).

Unfortunately, reliable and widespread success of RNAi gene knockdown is limited, especially in some insect taxa (Scott *et al*., 2013; Yu *et al*., 2013; Wynant *et al*., 2014). Numerous factors influence RNAi's performance including the inherent characteristics of the animal tested, the tissue target, target gene and the delivery methods employed (Scott *et al*., 2013; Yu *et al*., 2013; Wynant *et al*., 2014). To this end, more research is desperately needed to troubleshoot different RNAi delivery methods, and to tease apart which mechanisms ultimately determine RNAi success and failure in different nonmodel systems.

Within the insect order Hemiptera, there are several key economic agricultural pests with sequenced genomes in which RNAi technology has had varied success. Aphids, in particular, represent a group of hemipteran pests in which there is a wealth of genomic information making them key candidates in functional genomic studies. Many critical RNAi genes are present in aphids, suggesting that this group of insects would be ideal for RNAi gene knockdown (eg Jaubert-Possamai *et al*., 2010; Ortiz-Rivas *et al*., 2012; Bansal & Michel, 2013). However, in aphids there has been a large amount of variation in the success of using RNAi technology for gene knockdown. For example, plant-mediated RNAi within the green peach aphid, *Myzus persicae* (Sulzer), has high efficacy (Pitino *et al*., 2011; Guo *et al*., 2014; Coleman *et al*., 2015). By contrast, in the pea aphid, *Acyrthosiphon pisum* (Harris), plant-mediated RNAi is more difficult to develop, because *Ac. pisum*'s host plants are not easily transformable. Instead, siRNA or dsRNA have been delivered orally via artificial diets or by injection into the aphid. However, inconsistent and modest RNAi gene knockdown has been reported from these two delivery methods in *Ac. pisum* (Jaubert-Possamai *et al*., 2007; Shakesby *et al*., 2009; Mutti *et al*., 2006; Mao & Zeng, 2012; Christiaens *et al*., 2014). Christiaens *et al*. (2014) reported that dsRNAs delivered either orally or by injection are rapidly degraded in both aphid salivary secretions and haemolymph (insect blood) by putative RNAses, which may help explain why RNAi may not work efficiently in *Ac. pisum*.

A novel approach that has been successfully used in a few insect systems is the delivery of siRNA and dsRNA using nanoparticle carriers (Zhang *et al*., 2010; Li-Byarlay *et al*., 2013; Das *et al*., 2015). A recent study in the honeybee demonstrated successful gene knockdown of the targeted gene by aerosolizing a siRNA-perfluorocarbon nanoparticle emulsion through the honeybee's spiracles (Li-Byarlay *et al*., 2013). Nanoparticle carriers, such as the perfluorocarbon nanoparticle emulsion used by Li-Byarlay *et al*. (2013) facilitate the transfer of nucleic acids across biological barriers by enhancing their cellular uptake and by increasing the stability of dsRNA and siRNA (Ross *et al*., 2015).

In the current study, we determined the rate of dsRNA degradation after aerosolized nanoparticle delivery compared with microinjection in *Ac. pisum*. By aerosolizing the dsRNA with nanoparticles, we hoped to limit the dsRNA degradation that results from exposure to the haemolymph as shown in Christiaens *et al*. (2014) by directly targeting cellular tissues via tracheoles. Using similar methods and time points as Christiaens *et al*. (2014) we tested if double-stranded green fluorescent protein (dsGFP) RNA degrades rapidly inside of the aphid when using the aerosolized nanoparticle delivery method compared to microinjection.

We also investigated the efficacy of the aerosolized siRNA–nanoparticle technique in three aphid species: the pea aphid, *Ac. pisum*, the soybean aphid, *Aphis glycines* (Matsumura), and the greenbug, *Schizaphis graminum* (Rondani). We first determined if we could knock down the gene *carotene dehydrogenase* (*tor*; LOC100169245). In *Ac. pisum*, *tor* is a carotenoid gene that results in pink aphid pigmentation through torulene production (Moran & Jarvik, 2010). This gene was chosen because if *tor* is successfully knocked down, the aphid cuticle is expected to display a green-yellow phenotype, compared to the native pink colour displayed by our control *Ac. pisum* line. *Ap. glycines* and *S. graminum* were not included in the *tor* knockdown experiments because *S. graminum* does not produce torulene (Nováková & Moran, 2012) and the green-yellow pigmentation of *Ap. glycines* indicates a similar carotenoid profile to *S. graminum*.

For the second knockdown experiment, the gene *branched-chain amino acid transaminase* (*bcat*; LOC100167587) was selected because it is encoded in all aphid species in this study. Moreover, aphid-encoded *bcat* is hypothesized to be important in the regulation of the terminal step of the branched-chain amino acid biosynthesis pathways [eg leucine (Leu), valine (Val) and isoleucine (Ile); Wilson *et al*., 2010; Hansen & Moran, 2011; Poliakov *et al*., 2011], because it is not encoded within the aphid's bacterial symbiont's (*Buchnera*) genome (Shigenobu *et al*., 2000). We predict that when *bcat* is successfully knocked down there will be a reduction in the biosynthesis of these essential amino acids Leu, Val and Ile. *Bcat* is also important for the degradation of Leu, Val and Ile in other body tissues. We expect that a reduction in the biosynthesis and recycling of these branched-chain amino acids will result in lower aphid body mass compared to control aphids, because of the dysregulation of these essential nutrients.

In this study, we determined that *in vivo* dsRNA degradation was minimal when we delivered it via aerosolization, especially with the nanoparticle emulsion. Modest

gene knockdown of *tor* in *Ac. pisum* was observed at lower concentrations of siRNA, but no phenotypic effect was observed. We did not achieve successful knockdown of *bcat* in *Ac. pisum* or *S. graminum*. However, in *Ap. glycines*, we observed that aerosolized siRNA–nanoparticles significantly knocked down the target gene, *bcat*, and that there was a corresponding reduction in adult mass when compared to the control treatment. Overall, our findings indicate that aerosolized siRNA–nanoparticle delivery is a promising new high-throughput method of targeted gene knockdown in some aphid species. This method marks an improvement over traditional RNAi delivery methods (microinjection and artificial diets) because it can be widely applied to various non-model organisms in which the traditional RNAi delivery methods are not feasible or easily applied.

Results

GFP dsRNA degradation trials

Degradation of dsRNA *in vivo* was tested using two different dsGFP RNA delivery techniques: (1) direct microinjection and (2) a non-invasive aerosolization of siRNA–nanoparticle complexes using a nebulizer/compressor. For these experiments we used the *Ac. pisum* strain, LSR1.

dsRNA degradation trials: microinjection. Our results indicate that in dsGFP RNA-microinjected aphids, dsGFP RNA does not significantly degrade over time *in vivo*. However, there was a trend of degradation [analysis of variance (ANOVA); $F = 2.28$, df $= 4$, $P = 0.093$; Fig. 1].

dsRNA degradation trials: non-invasive aerosolization of nanoparticle complexes. In contrast to injection treatments, levels of intact dsGFP RNA did significantly decrease over time for aphids that were treated with aerosolized dsGFP RNA alone without nanoparticles (Treatment 1, $F = 3.10$, df $= 4$, $P = 0.04$; Fig. 2). Levels of intact dsGFP RNA significantly decreased 5 h after the aerosolized dsGFP RNA treatment. However, 30 min, 1 and 2 h treatments are not significantly different at $\alpha \leq 0.05$ (Fig. 2, Table 1). In comparison, for both aerosolized dsGFP RNA treatments with nanoparticles (Treatments 2 and 3), which varied fivefold in dsGFP RNA concentration, dsGFP RNA did not significantly degrade over time (ANOVA; $F = 2.07$, df $= 4$, $P = 0.12$; $F = 0.66$, df $= 4$, $P = 0.63$, respectively, Fig. 2). Overall, these results indicate that dsGFP RNA delivered either through microinjection or aerosolization, especially with nanoparticles, is highly stable inside *Ac. pisum* (LSR1) over time.

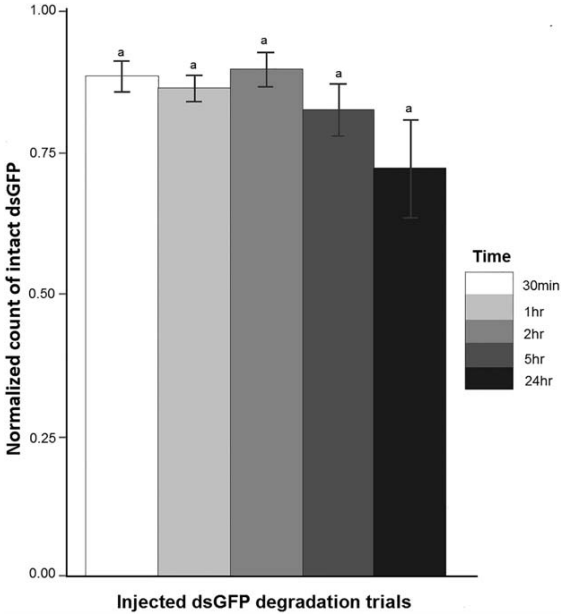


Figure 1. Normalized count of intact double-stranded green fluorescent protein (dsGFP) RNA observed in microinjected double-stranded RNA degradation trials. The counts of GFP transcripts were standardized to an aphid housekeeping gene (*elongation factor 1-alpha*). Error bars indicate ± 2 SD from the mean. Bars with different letters above represent significant treatment differences within a trial at $P < 0.05$ based on post-hoc comparisons (least significant difference tests). Non-GFP treated aphids served as negative controls and dsGFP RNA was not present as expected (bars not shown).

siRNA–nanoparticle aerosolization trials

During each experiment, there were three treatments: (1) target siRNA and nanoparticle emulsion (siRNA + nanoparticles), (2) control siRNA and nanoparticle emulsion (control siRNA + nanoparticles) and (3) siRNA only. For gene knockdown experiments, the sample size for each treatment was six aphids. The control siRNA was a randomized siRNA sequence that had no predicted target within the aphid.

Knockdown of the tor *gene.* For the *tor* gene knockdown experiments, we used the *Ac. pisum* strain LSR1. For trial TOR-1, which tested the siRNA concentration of 100 nM, we found a significant decrease in *tor* gene expression for the *tor* siRNA + nanoparticle experimental treatment compared to the control treatment (*t*-value $= -2.08$, $P = 0.03$; Fig. 3A). This *tor* siRNA + nanoparticle treatment resulted in a $\sim$12% knockdown compared to the control. Significant knockdown was not found for the siRNA-only treatment compared to the control (*t*-value $= -1.69$, $P = 0.06$; Fig. 3A), indicating that nanoparticles are important for *tor* knockdown in *Ac. pisum* (LSR1). For trial TOR-2, which used a twofold higher concentration of *tor*-siRNA, a significant

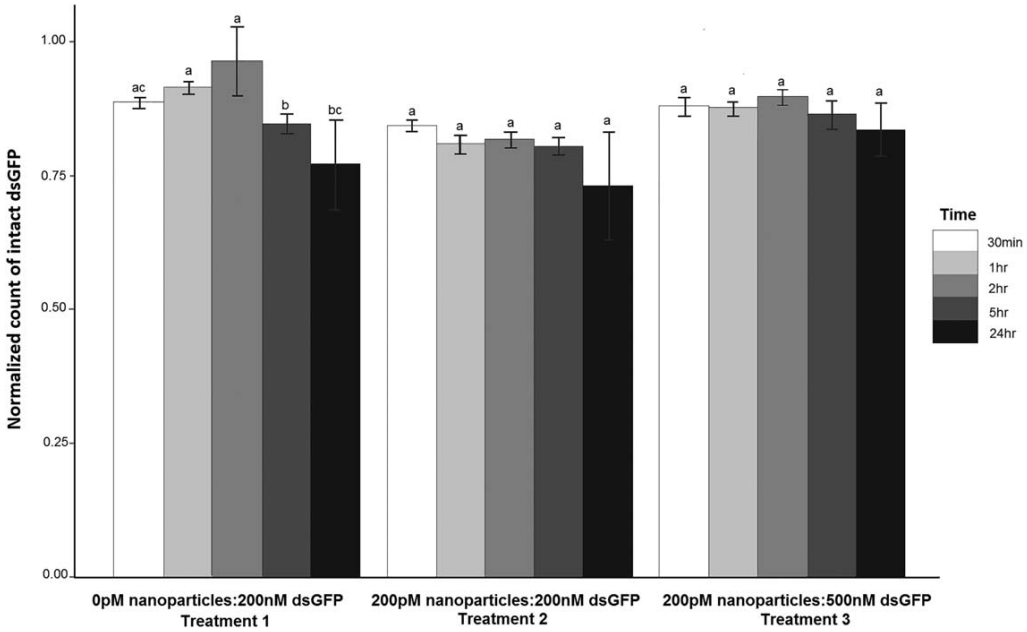


Figure 2. Normalized count of intact double-stranded green fluorescent protein (dsGFP) RNA observed in aerosolization double-stranded RNA degradation trials. The counts of GFP transcripts were standardized to an aphid housekeeping gene (*elongation factor 1-alpha*). Error bars indicate ± 2 SD from the mean. Bars with different letters above represent significant treatment differences within a trial at $P < 0.05$ based on post-hoc comparisons (least significant difference tests). Non-GFP treated aphids served as negative controls and dsGFP RNA was not present as expected (bars not shown).

difference in *tor* gene expression was observed in the *tor* siRNA + nanoparticles compared to the control (t-value = −2.90, $P \leq 0.01$) but the percent knockdown was less than in TOR-1 at ∼8% (Fig. 3B). Like the TOR-1 treatment, significant knockdown was not found for the siRNA-only treatment compared to the control (t-value = 0.64, $P = 0.64$; Fig. 3B), indicating that nanoparticles are important for *tor* knockdown in *Ac. pisum* (LSR1) at a siRNA concentration of 200 nM. When we increased the concentration of siRNA in trials TOR-3 (siRNA concentration 500 nM) and TOR-4 (siRNA concentration 1000 nM) in *Ac. pisum* (LSR1) (Fig. 3C, D, Table S3) we observed no significant gene knockdown.

Table 1. Pairwise comparison of time points (T) from Treatment 1 double-stranded RNA degradation aerosolization assays

	T1 (30min)	T2 (1 h)	T3 (2h)	T4 (5 h)	T5 (24 h)
T1 (0.5 h)	–	0.539	0.133	0.425	0.063
T2 (1 h)	–	–	0.344	0.165	**0.022**
T3 (2h)	–	–	–	**0.029**	**0.004**
T4 (5 h)	–	–	–	–	0.207
T5 (24 h)	–	–	–	–	–

Significantly different treatments are in bold.
Acyrthosiphon pisum aphids in trial Treatment 1 were aerosolized with 100 nM double-stranded green fluorescent protein (dsGFP) RNA without nanoparticles.
Least significant difference tests were conducted on normalized expression values for GFP standardized to an aphid housekeeping gene.

We also evaluated aphid pigment changes after all treatments prior to isolating RNA because *tor* is responsible for the aphid's pink pigmentation. Regardless of trial or treatment, we did not observe any visible change in aphid pigmentation, compared to the corresponding trial's control treatment in *Ac. pisum* (LSR1).

Knockdown of the bcat *gene.* siRNA concentrations for the *bcat* experiments were optimized based on results from the first set of experiments with *tor*. Consequently, for *bcat* knockdown we tested the concentrations of siRNA at 100 nM (BCAT-100) and 200 nM (BCAT-200). We found no significant difference in *bcat* expression in *S. graminum* and the two strains of *Ac. pisum*, between control and experimental treatments [(1) siRNA + nanoparticles, (2) control siRNA + nanoparticles and (3) siRNA only (Fig. 4 A, B, Tables S4, S5)]. For these aphid species, we also observed no significant change in aphid mass between control and experimental treatments (Fig. 4C, D, Tables S4, S5).

In contrast to *Ac. pisum* (5A and LSR1) and *S. graminum*, in *Ap. glycines* we observed both a significant gene knockdown and an effect on aphid mass for the aerosolized target siRNA nanoparticle trials (BCAT-200). Specifically, in the BCAT-200 *Ap. glycines* trials we observed a ∼30% knockdown in the siRNA + nanoparticle treatment (t-value = 3.67 $P \leq 0.001$; Fig. 5B) and ∼19% in

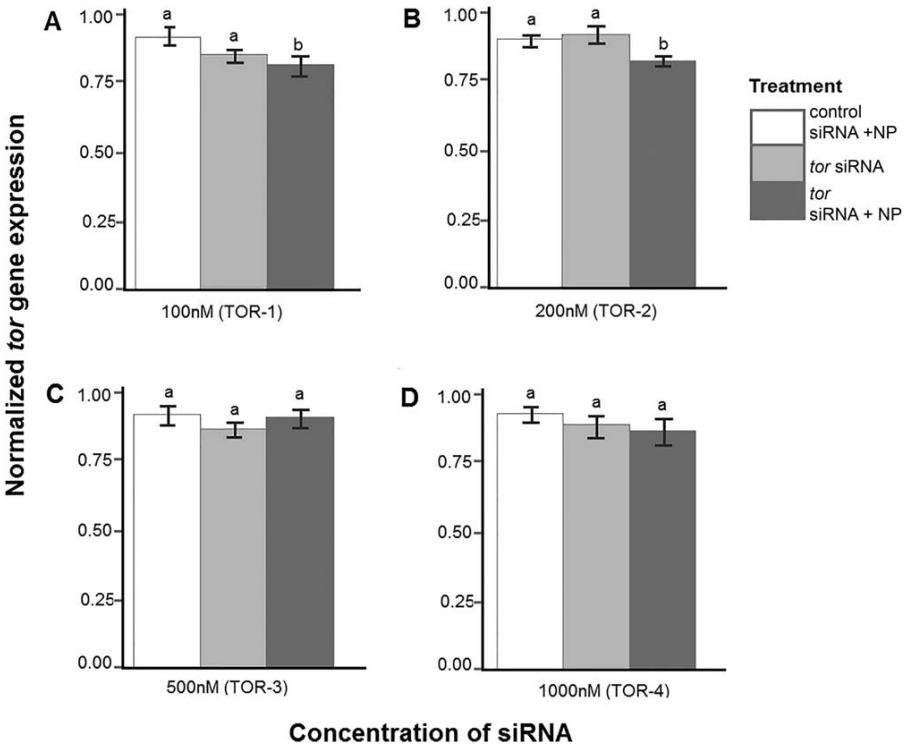


Figure 3. Normalized *carotene dehydrogenase* (*tor*) gene expression levels for the aerosolized siRNA-nanoparticle (siRNA-NP) trials (TOR-1, TOR-2, TOR-3 and TOR-4) for *Acyrthosiphon pisum* (LSR1). All expression values for *tor* were standardized to an aphid housekeeping gene (*elongation factor 1-alpha*). See Table 1 for sample sizes. Error bars indicate ± SEM. Bars with different letters above represent significant differences within a trial at $P < 0.05$. Controls for all trials were aerosolized NP–control siRNA treatments.

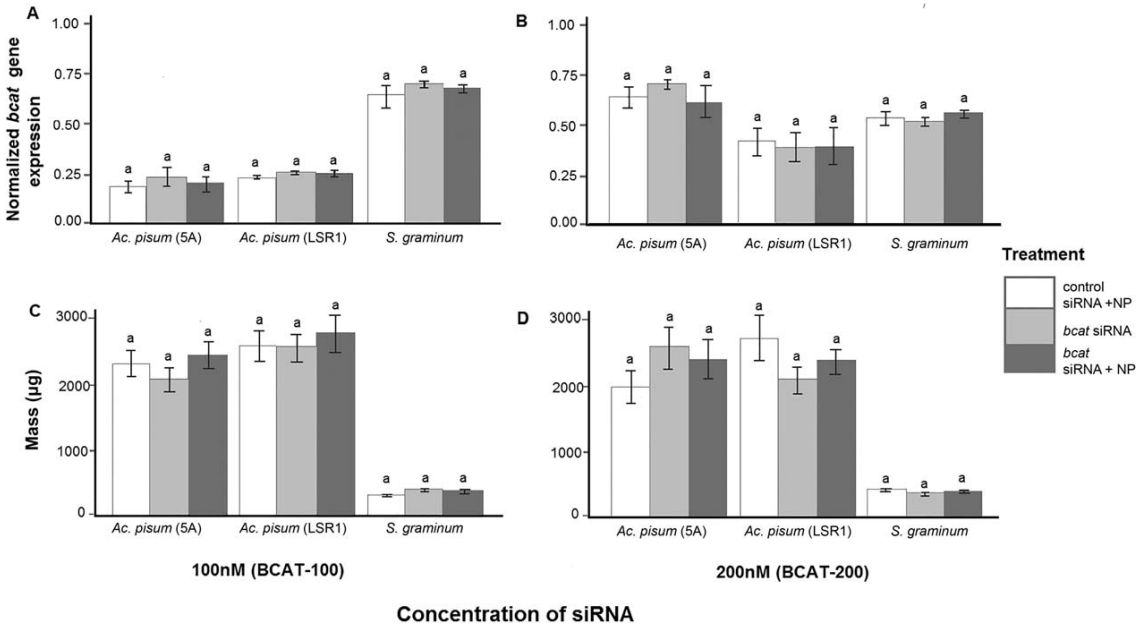


Figure 4. Normalized *branched-chain amino acid transaminase (bcat)* gene expression levels and aphid mass for the aerosolized siRNA-nanoparticle (siRNA-NP) trials (BCAT-100 and BCAT-200) for *Acyrthosiphon pisum* (5A, LSR1) and *Schizaphis graminum*. See Table 2 for gene expression experiment sample sizes and Tables S4 and S5 for aphid mass experiment sample sizes. All expression values for *bcat* were standardized to an aphid housekeeping gene (*elongation factor 1-alpha*). Error bars indicate ± SEM from the mean. Bars with different letters above represent significant differences within a trial at $P < 0.05$.

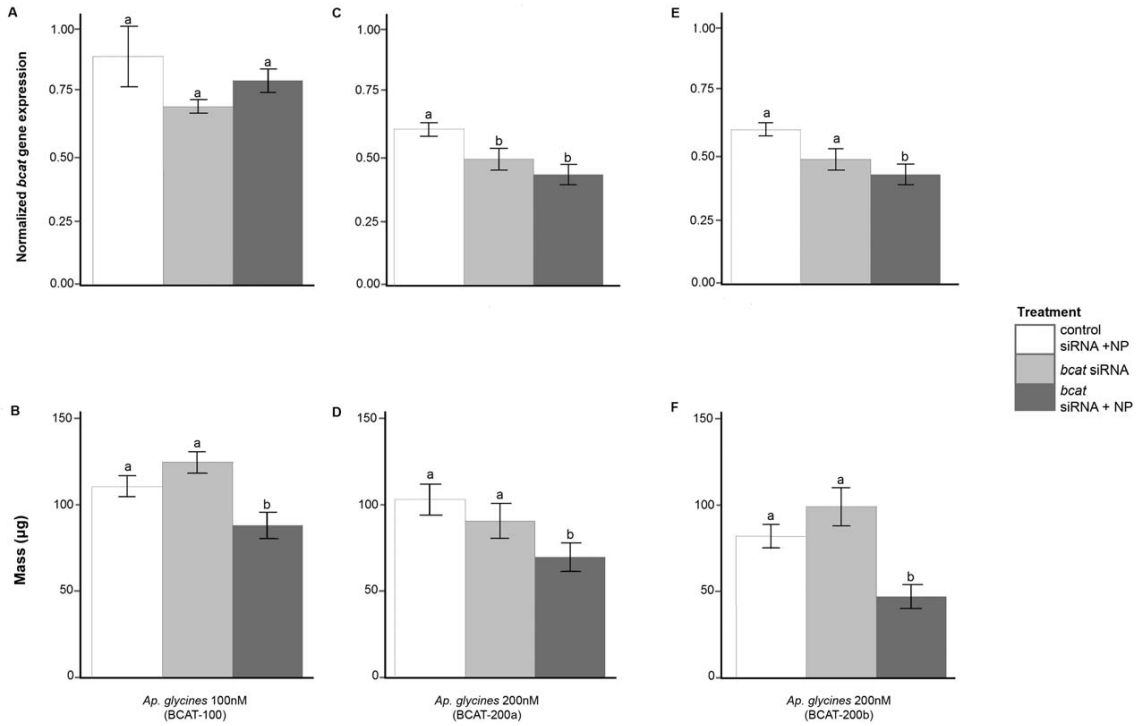


Figure 5. Normalized *branched-chain amino acid transaminase* (*bcat*) gene expression levels and aphid mass for the aerosolized siRNA- nanoparticle trials for *Ap. glycines*. All expression values for *bcat* were standardized to the housekeeping gene *ribosomal protein S9 (RSP9)*. BCAT- 200a and BCAT-200b represent two independent trials: See Table 2 for gene expression experiment sample sizes and Tables S4 and S5 for aphid mass experiment sample sizes. Error bars indicate ± SEM. Bars with different letters above represent significant differences within a trial at $P < 0.05$.

the siRNA only treatment (*t*-value = 2.35, $P = 0.02$; Fig. 5B) when compared with the control. This gene knockdown was accompanied by a decrease in mass in the siRNA + nanoparticle trial when compared with the control (*t*-value = 2.69, $P = 0.01$; Fig. 5E). A significant decrease in mass was not found for the siRNA-only treatment compared with the control (*t*-value = 0.91, $P = 0.19$; Fig. 5E). These results strongly suggest that nanoparticles are important for the successful delivery of siRNAs in *Ap. glycines*. We were able to replicate a similar level of significant gene knockdown in the siRNA + nanoparticle treatment compared with the control in an independent trial and observed a ∼30% knockdown (*t*-value = 1.92, $P = 0.05$; Fig. 5C). In the replicate trial, we also observed a reduction in overall aphid mass in the *bcat* siRNA + nanoparticles treatment when compared with the control treatment (*t*-value = 3.62, $P \leq 0.001$; Fig. 5F), once more indicating that nanoparticles are important for the successful delivery of siRNAs in *Ap. glycines*.

In the *Ap. glycines* BCAT-100 trial we observed a significant decrease in aphid mass in the *bcat* siRNA treatment + nanoparticle treatment compared with the control (*t*-value = 2.13, $P = 0.02$; Fig. 5A). However, this was not reflected in the relative gene expression results. Based on the reverse transcriptase-quantitative PCR (RT-qPCR) results we observed no significant gene knockdown in the *bcat* siRNA + nanoparticle treatment (*t*-value = −0.75, $P = 0.24$; Fig. 5B), and no significant knockdown in the *bcat* siRNA-only treatment compared with the control treatment (*t*-value = −1.62, $P = 0.08$; Fig. 5B). This suggests that lower concentrations of siRNA do not result in significant knockdown of *bcat*.

Nanoparticles have no effect on aphid mass or mortality. In the trials with *Ap. glycines* and *S. graminum*, we observed a greater level of aphid mortality after the aerosolized spray treatments when compared with treatments conducted on *Ac. pisum*. To understand aerosolization treatment effects on these two species we exposed *Ap. glycines* and *S. graminum* aphids to two additional experimental treatments: water + 200 pM nanoparticles, and water only.

Overall, aerosolized water with or without nanoparticles has no effect on aphid mass in *Ap. glycines* or *S. graminum* (*S. graminum:* *t*-value = 0.97, $P = 0.17$; *Ap. glycines:* *t*-value = 0.44, $P = 0.67$; Fig. S2A). In both *S. graminum* and *Ap. glycines* we observed no significant differences in mortality when aphids were exposed to

aerosolized water with or without nanoparticles (*S. graminum*: z-score = 0.40, $P = 0.35$; *Ap. glycines*: z-score = 0.64, $P = 0.26$; Fig. S2B). It is important to note that the mortality that we observed when aphids were exposed to aerosolized water with nanoparticles was not significantly different from background mortality (*S. graminum*: z-score = 0.40, $P = 0.35$; *Ap. glycines*: z-score = 0.64, $P = 0.2$; Fig. S2B). Background mortality observed in *Ap. glycines* was similar to what was observed by McCornack *et al.* (2004) at similar temperatures and life stages.

Discussion

Results from our study suggest that a new RNAi delivery technique used in aphids that aerosolizes siRNA–nanoparticles into insect spiracles can result in successful, targeted gene knockdown in some aphid species depending on siRNA concentrations and gene targets. In our study, gene knockdown was observed in *Ap. glycines* and to a lesser extent in *Ac. pisum* (LSR1) when aerosolized siRNAs and nanoparticles were administered to aphids compared with aerosolized siRNAs without nanoparticles, or a control treatment consisting of control siRNA + nanoparticles (Figs 2A, B, 4, 5). These results indicate that siRNAs in combination with nanoparticles are required for successful target gene knockdown when using this particular mode of aerosolized delivery into aphids.

Variable success in RNAi knockdown has been a major obstacle curtailing widespread implementation of RNAi in insects and other organisms. In this study, we also observed lineage-specific responses to RNAi knockdown. We attempted to knock down gene expression of two gene targets in *Ac. pisum*, *bcat* and *tor*, and were able to achieve only modest gene knockdown of *tor* and no gene knockdown of *bcat*. Within the *tor* trials, we observed a modest gene knockdown at the lower concentrations tested: 100 and 200 nM (trials TOR-1 and TOR-2, respectively, Fig. 3A,B). However, when we increased siRNA concentrations we did not see a concomitant increase in target gene knockdown (trials TOR-3 and TOR-4). By contrast, the higher siRNA concentration of 1000 nM achieved efficient knockdown in the honeybee, where a 30% reduction in the expression of DNA methyltransferase 3 *(dnmt3)* was observed using aerosolized siRNA–nanoparticles (Li-Byarlay *et al.*, 2013). A possible explanation for this variation may be that at higher concentrations the charge of the siRNA–nanoparticle complex is altered, so that cellular uptake into aphid cells is disrupted. Both the current study and the honeybee study (Li-Byarlay et al., 2013) used the same nanoparticle emulsion. This nanoparticle emulsion utilizes the lipid–raft transport system to traffic molecules into the cell and requires a net positive charge (Kaneda *et al.*, 2010). This positive charge is achieved when the siRNAs bind to the nanoparticle; however, different siRNA concentrations may result in different net charges of the complex so that the rate of uptake is negatively impacted in different tissue types (He *et al.*, 2010). The *tor* carotenoid gene is responsible for the pink pigmentation in *Ac. pisum* through torulene production (Moran & Jarvik, 2010). Although we were able to measure gene knockdown in the TOR-1 and TOR-2 trials we did not observe a colour change in aphid pigmentation compared with the corresponding trial's control treatments. However, even if *tor* is knocked down a change in pigmentation may not occur because carotenoid proteins are extremely stable in many organisms after their production (Yahia and Ornelas-Paz, 2010).

We were able to observe successful knockdown of *bcat* in *Ap. glycines*, but not in *S. graminum* or *Ac. pisum*. The aphid gene *bcat* is hypothesized to be important in the production of branched-chain amino acids in the aphid by utilizing metabolic intermediates produced by the aphid symbiont's (*Buchnera*) branched-chain amino acid pathways (Wilson *et al.*, 2010; Hansen & Moran, 2011; Poliakov *et al.*, 2011). Therefore, by targeting this aphid gene in bacteriocytes (specialized aphid cells that contain the aphid symbiont *Buchnera*), we would be disrupting a key gene that is involved in complementing *Buchnera*'s essential amino acid pathways. We hypothesize that when *bcat* is successfully knocked down in aphids, a reduction in the biosynthesis of these essential amino acids (Leu, Val and Ile) will result. In consequence, the predicted reduction of these essential amino acids will result in aphid starvation, reducing aphid body mass compared with the wild-type. In the *Ap. glycines* BCAT-200 trials, we were able to observe not only successful gene knockdown of *bcat* in the aerosolized siRNA and nanoparticle treatment but also a concomitant change in aphid mass compared with the control, which matched our predictions (Fig. 5). It is important to note that this effect on aphid mass can also be associated with the disruption of branched-chain amino acid degradation in other body tissues as well. We were able to determine that any changes in aphid mass were not associated with nanoparticles or water, as our water only, water + nanoparticles (Fig. S2A) and our control siRNA + nanoparticle treatments did not result in a body mass reduction.

The successful gene knockdown observed in this study may be due in part to the fact that nanoparticles bound to siRNA molecules tend to be more stable and more likely to undergo cellular uptake into cells (Ross *et al.*, 2015). However, we found that dsRNAs, which are double stranded like siRNAs, do not degrade rapidly in the pea aphid *in vivo* using either microinjection or

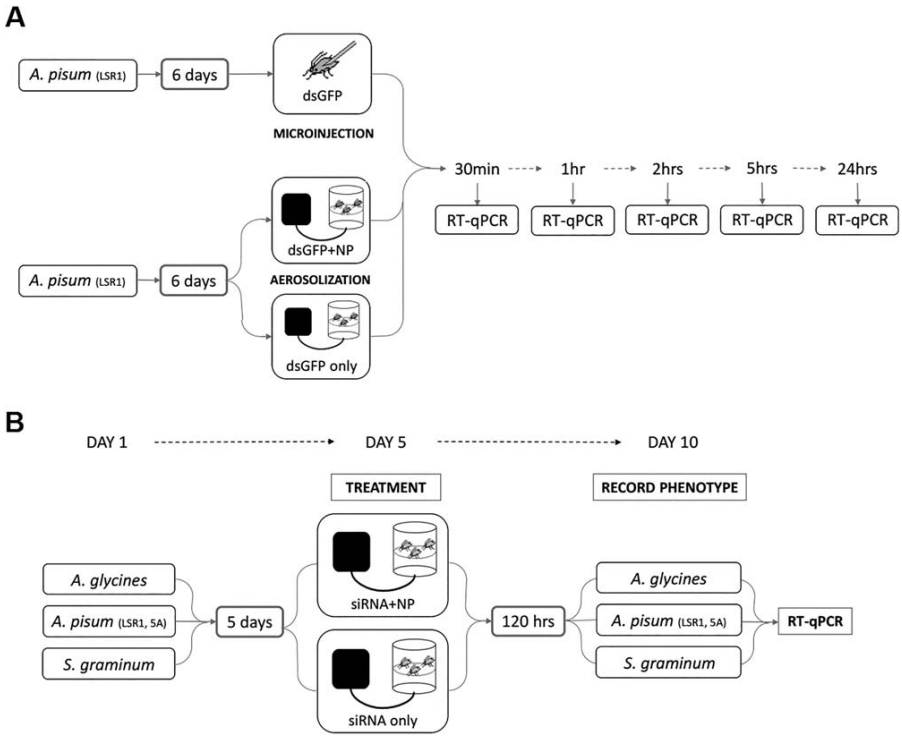


Figure 6. Overview of double-stranded RNA (dsRNA) degradation and short interfering RNA (siRNA)–nanoparticle emulsion aerosolization experiment methodology. (A) dsRNA degradation experiments: 6-day-old aphids were exposed to dsRNA either via microinjection or aerosolization. dsRNA degradation was calculated via reverse transcriptase-quantitative PCR (RT-qPCR) at five time points after the treatment. (B) siRNA–nanoparticle emulsion aerosolization experiments: 5-day-old aphids were exposed to siRNAs with or without the nanoparticle emulsion. The control treatment consisted of a randomized siRNAs + nanoparticles. After 120 h any change in phenotype was recorded (colour or mass) and gene expression levels were determined via RT-qPCR. Abbreviations: *A. glycines*, *Aphis glycines*; *A. pisum*, *Acyrthosiphon pisum*; dsGFP, double stranded green fluorescent protein RNA; NP, nanoparticle; *S. graminum*, *Schizaphis graminum*.

aerosolization, with or without nanoparticles as a carrier (Figs (1 and 2)). These results are in contrast to those of Christiaens *et al*. (2014). Our dsRNA degradation results may differ from Christiaens *et al*. (2014) because we used a different pea aphid strain, GFP plasmid and/or aphid life stage. Regardless, no rapid degradation of dsRNA was observed in our *Ac. pisum* strain via either microinjection or aerosolized dsRNA–nanoparticle delivery. Therefore, we do not expect RNA degradation of dsRNA to impact RNAi in this new mode of delivery using the aerosolization of nanoparticles. In sum, our results suggest that *in vivo* siRNA degradation may not be an important factor affecting the success of RNAi for this delivery method or aphid line. In general, the results from our study indicate that aerosolized siRNA–nanoparticle complexes may be more efficient at undergoing cellular uptake compared with siRNA molecules alone, suggesting that nanoparticles are key for increasing the efficiency of RNAi using this delivery technique.

The collection time after the siRNA treatment seems to play an important role in whether or not gene knockdown is observed. Previous RNAi studies conducted on pea aphids using microinjection or ingestion observed gene knockdown that ranged from 27–46% on day five or seven post-RNAi treatment, regardless of the delivery method (Mutti *et al*., 2006; Jaubert-Possamai *et al*., 2007; Shakesby *et al*., 2009; Mao & Zeng, 2012). We observed a similar trend in *Ap. glycines* in which no gene knockdown was observed when aphids were screened earlier than 5 days post-siRNA exposure (Fig. S3).

Aphis glycines was the smallest aphid that we tested in this study [*Ap. glycines* is ~3 times smaller in body length and ~27 times smaller in mass compared to *Ac. pisum* (Blackman & Eastop, 2006; Figs 5C, D, 6C, D, S2A)], and its small size may have increased the efficacy of gene knockdown given our selected experimental concentrations and volumes. For example, it may be possible that we were able to expose *Ap. glycines*' internal tissues to a greater volume of siRNA–nanoparticles because this species has a larger surface-area-to-volume ratio when compared with *Ac. pisum* and *S. graminum*.

Overall, this new mode of aerosolized siRNA–nanoparticle delivery is a promising high-throughput and non-invasive RNAi technique that should be optimized further and

used to investigate gene functions in aphids and other insect systems. Once optimized, this method would be an improvement over traditional RNAi delivery methods (microinjection and artificial diets) because it can be widely applied in systems where the traditional RNAi delivery methods are not feasible or easily applied. For example, to implement the feeding technique an artificial diet must be created; however, one may not be available for all insect systems. The development of an artificial diet is not always easy, and for the soybean aphid a successful artificial diet has not been fully developed (R. Bansal, pers. comm.). Alternatively, a transgenic plant must be made, which is time consuming and much easier to implement in some model plants systems compared with others. Microinjection, although successful for various insect species, has the disadvantage of being highly invasive and difficult for small insects such as the soybean aphid. It also has the disadvantage of being a relatively slow technique and requires expensive, specialized equipment and experience. Via aerosolization, ∼20 aphids can be exposed to a RNAi treatment in 5 min and only requires a nebulizer.

In this study, we observed modest to moderate levels of gene knockdown; however, it is important to note that these levels were for whole body samples. The efficiency of gene knockdown may be higher in specific tissue types using this technique, as was demonstrated by Li-Byarlay *et al*. (2013). This highlights the need for future studies to determine if this new method of gene silencing in aphids can be optimized further, and which aphid cells and tissues successfully uptake siRNA–nanoparticle complexes. Many aphid species are major economic pests and in the long term this technology may potentially be used in their management. Results generated from such studies will ultimately augment research productivity and progress in the exploding field of functional genomics.

Experimental procedures

Aphid colonies

The *Ac. pisum* strains 5A and LSR1 were reared on broad bean (*Vicia faba*), *S. graminum* was reared on barley (*Hordeum vulgare* L.) and *Ap. glycines* biotype 1 (Kim *et al*., 2008; Cooper *et al*., 2015) was reared on susceptible soybean seedlings (*Glycine max*, cultivar Williams82). *Ac. pisum* and *S. graminum* aphid colonies were maintained at 20 °C with a 16 h light: 8 h dark cycle (Hansen & Moran, 2012). *Ap. glycines* was maintained at 24 °C with a 14 h light: 10 h dark cycle, which was similar to the conditions used by Bansal *et al*. (2014). Before each experiment, even-age cohorts were established. To establish these cohorts, 25–30 adults were placed on a plant; within 24 h all adults were removed from the plant leaving the 1-day-old nymphs to develop on the plant.

GFP dsRNA degradation trials

Degradation of dsRNA *in vivo* was tested using two different dsGFP RNA delivery techniques: (1) direct microinjection and (2) a non-invasive aerosolization of siRNA–nanoparticle complexes using a nebulizer/compressor (Probasics, PMI, Marlboro, NJ, USA). For both the direct microinjection and aerosolization assays the degradation process was measured at five time points after the dsGFP RNA delivery (30 min, 1 h, 2 h, 5 h and 24 h; Table S2). For both techniques, 30 six-day-old LSR1 *Ac. pisum* aphids (fourth instar–early adult) were exposed to each treatment. At each time point, six individuals were collected, snap-frozen at −80 °C on dry ice, and preserved in RNA*later*-ICE (Ambion Life Technologies Corporation, Grand Island, NY, USA) (see Fig. 6A for experimental design).

dsRNA preparation. dsRNA was used in this experiment because unlike siRNA, it is long enough to test for degradation using RT-qPCR. Synthesized dsGFP RNA was used for degradation assays on aphids *in vivo*. First, DNA template of 424 bp of superfold GFP (sfGFP) was amplified from the pZEMB8 plasmid that we kindly received as a gift from the C.K. Vanderpool lab in the Department of Microbiology at the University of Illinois, Urbana-Champaign. Specific primers (dssfGFP-F/R), containing the T7 promotor sequence at each 5′ end were used for the initial PCR reaction (Table S1). dsRNA for sfGFP was synthesized *in vitro* using a Megascript RNAi Kit (Ambion Life Technologies Corporation) from 1 μg amplified PCR template. To maximize dsRNA yield, the composition of the binding mix in the purification step was altered and three volumes of unpurified dsRNA were used. The concentration of dsGFP RNA was determined using a NanoDrop Lite Spectrophotometer (Thermo-Fisher Scientific, Wilmington, DE, USA). Furthermore, dsGFP RNA stability in the supplied elution buffer (Ambion Life Technologies Corporation) and in water was assessed after 24 h using 1% AQ29 agarose gel electrophoresis per the manufacturer's protocol (data not shown). Based on this quality check, synthesized dsGFP appeared to be of high quality and in the correct size range.

dsRNA degradation trials: microinjection. For the microinjection assay, ∼40 ng dsGFP RNA was administered into each individual using a Discovery V8 stereo-microscope (Zeiss, Oberkochen, Germany) with the Micro 4 MicroSyringe pump controller micro-injector (World Precision Instruments, Sarasota, FL, USA) using a 34 GA, 1.97′, 30° microinjection needle (Hamilton, Reno, NV, USA).

dsRNA degradation trials: non-invasive aerosolization of nanoparticle complexes. The nanoparticle emulsion used in this study was designed by the Wickline laboratory (Kaneda *et al*., 2010) and kindly provided to us. The nanoparticle emulsion used was the same as the one used by Li-Byarlay *et al.* (2013) to successfully knock down gene expression in the honeybee. Li-Byarlay *et al*. (2013) also demonstrated that the siRNA–nanoparticle complexes successfully penetrate the spiracles on the thorax and abdomen and travel through the tracheal respiratory system. For the dsRNA degradation aerosolization assays, aphids were administered with 3 ml aerosolization mixture, which contained 200 pM of the nanoparticle

Table 2. Aerosolized short interfering RNA (siRNA)–nanoparticles trials and treatments testing the knockdown of the *carotene dehydrogenase* (*tor*) gene in *Acyrthosiphon pisum* (LSR1)

Trial	NP concentration (pM)	siRNA concentration (nM)	Aphid age during spray treatment (days)	Spray treatment (sample size for RT-qPCR)	Sample collection time after spray (h)	Cuticle colour	Aphid species (line)
TOR-1	200	100	5	ctrl siRNA + NP (n = 6)	120	Pink	*Ac. pisum*
				tor siRNA + NP (n = 6)		Pink	*Ac. pisum*
				tor siRNA (n = 6)		Pink	*Ac. pisum*
TOR-2	200	200	5	ctrl siRNA + NP (n = 6)	120	Pink	*Ac. pisum*
				tor siRNA + NP (n = 6)		Pink	*Ac. pisum*
				tor siRNA (n = 6)		Pink	*Ac. pisum*
TOR-3	200	500	5	ctrl siRNA + NP (n = 6)	120	Pink	*Ac. pisum*
				tor siRNA + NP (n = 6)		Pink	*Ac. pisum*
				tor siRNA (n = 6)		Pink	*Ac. pisum*
TOR-4	200	1000	5	ctrl siRNA + NP (n = 6)		Pink	*Ac. pisum*
				tor siRNA + NP (n = 6)	120	Pink	*Ac. pisum*
				tor siRNA (n = 6)		Pink	*Ac. pisum*

ctrl, control; NP, nanoparticle; RT-qPCR, reverse transcriptase-quantitative PCR.

emulsion, varying in concentrations of dsRNA and water. Treatment 1 contained 100 nM dsGFP RNA only, Treatment 2 contained 100 nM dsGFP RNA + 200 pM nanoparticles, and Treatment 3 contained 500 nM dsGFP RNA + 200 pM nanoparticles.

siRNA–nanoparticle emulsion aerosolization trials

For all experiments, 20–30 aphids were placed in an enclosed container and a 3-ml siRNA solution was aerosolized using a nebulizer/compressor (Probasics, PMI). The 3 ml solution contained 200 pM of the nanoparticle emulsion, varying concentrations of siRNA and water. During each experiment, there were three treatments: (1) target siRNA and nanoparticle emulsion (siRNA + nanoparticles), (2) control siRNA and nanoparticle emulsion (control siRNA + nanoparticles) and (3) siRNA only. The control siRNA was a randomized siRNA sequence that had no predicted target within the aphid. The concentration of siRNA varied between trials; however, the concentration of nanoparticles remained constant (200 pM) for all trials. This was the same concentration that was used in the degradation trials. All siRNAs were ordered through Sigma (St Louis, MO, USA). Five-day-old aphids were exposed to the aerosolized siRNA mixture for 5 min, after which aphids were moved back to their host plant until they were collected. Aphids were then collected 5 days after spray treatment, snap-frozen on dry ice and preserved in RNA*later*-ICE (Ambion Life Technologies Corporation) (see Fig. 6B for experimental design). These time points and aphid ages were selected based on preliminary experiments (data not shown) and time points used in other aphid RNAi experiments (eg Mutti *et al.*, 2006; Jaubert-Possamai *et al.*, 2007; Shakesby *et al.*, 2009; Mao & Zeng, 2012).

Knockdown of the tor *gene.* A control siRNA was designed by randomizing the *tor* nucleotide sequence (Sigma). Control and target *tor* siRNA oligo sequences are given in Table S1. Four trials for the *tor* gene that varied in siRNA concentrations, hereafter referred to as TOR-1 to TOR-4, were performed in *Ac. pisum* (LSR1). Each trial had a specific siRNA concentration: TOR-1: 100 nM siRNA; TOR-2: 200 nM siRNA; TOR-3: 500 nM siRNA; and TOR-4: 1000 nM siRNA. Sample sizes for each trial are detailed in Table 2.

Knockdown of the bcat *gene.* For the second set of siRNA–nanoparticle aerosolization experiments, knockdown of *bcat* was tested on *Ac. pisum* (5A and LSR1), *Ap. glycines* and *S. graminum*. Species-specific *bcat* siRNAs were designed for each species. Control and target *bcat* siRNA oligonucleotide sequences are presented in Table S1. siRNA concentrations for the *bcat* experiments were optimized based on the results from the first set of experiments with *tor*. Consequently, for *bcat* knockdown we tested the concentrations of siRNA at 100 nM (BCAT-100) and 200 nM (BCAT-200) (Table 3). Aphid masses were recorded using a UMX2 microscale (Mettler Toledo, Columbus, OH, USA) (see Fig. 6B for experimental design).

After observing greater aphid mortality in the *Ap. glycines* and *S. graminum* trials when compared with *Ac. pisum*, we exposed *Ap. glycines* and *S. graminum* aphids to two additional experimental treatments: water + 200 pM nanoparticles, and water only. These experiments were to help us understand aerosolization treatment effects on these two species. We then recorded aphid mortality and aphid mass (Fig. S4). We also determined background mortality rates for these two species (Fig. S4B).

RNA extractions and RT-qPCR

To determine if there was a difference in dsGFP RNA abundance for dsRNA degradation and target gene mRNA abundance for the nanoparticle–siRNA aerosolization trials detailed above, total RNA was extracted from individual aphid whole bodies using a Qiagen RNeasy Mini kit (Qiagen, Valencia, CA, USA) with the Qiagen RNase-Free DNase treatment (Qiagen). Aphids that displayed low quality/quantity RNA (determined by spectrometry – NanoDrop lite, Thermo-Fisher Scientific), which prevented cDNA synthesis, were removed from the analysis.

For cDNA synthesis, an All-in-One cDNA Synthesis SuperMix kit (Biotool, Houston, TX, USA) was used with 200 ng total RNA for each sample following the manufacturer's protocols.

Table 3. Aerosolized short interfering RNA (siRNA)–nanoparticles trials testing the knockdown of the *branched-chain amino acid transaminase* (*bcat*) gene in three aphid species

Trial	NP concentration (pM)	siRNA concentration (nM)	Aphid age during spray treatment (days)	Spray treatment (sample size for RT-qPCR)	Sample collection time after spray (h)	Aphid species (line)
BCAT 100	200	100	5	ctrl siRNA + NP (n = 6)	120	*Acyrthosiphon pisum* (5A, LSR1), *Aphis glycines*, *Schizaphis graminum*
				bcat siRNA + NP (n = 6)		*Ac. pisum* (5A, LSR1), *Ap. glycines*, *S. graminum*
				bcat siRNA (n = 6)		*Ac. pisum* (5A, LSR1), *Ap. glycines*, *S. graminum*
BCAT 200	200	200	5	ctrl siRNA + NP (n = 6)	120	*Ac. pisum* (5A, LSR1), *Ap. glycines*, *S. graminum*
				bcat siRNA + NP (n = 6)		*Ac. pisum* (5A, LSR1), *Ap. glycines*, *S. graminum*
				bcat siRNA (n = 6)		*Ac. pisum* (5A, LSR1), *Ap. glycines*, *S. graminum*

ctrl, control; NP, nanoparticle.

RT-qPCR was conducted on each sample using the SYBR Fast Universal qPCR reagents (KAPA Biosystems, Woburn, MA, USA) and iTaq Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) on a Bio-Rad CFX96 Real-Time PCR Detection System. No aphid samples were pooled. All qPCR experiments were performed on individual aphids.

For the dsGFP RNA degradation trials, the presence of intact, nondegraded dsGFP was detected using RT-qPCR with the primers qsfGFP-F/R 9 (Table S1), which amplify a 102-bp region of our 424-bp dsRNA target. Untreated aphids served as negative controls for RT-qPCR detection of dsGFP because dsGFP is not endogenously expressed in the aphids. For control, non-GFP treated aphids, expression of the aphid housekeeping gene *elongation factor 1-alpha* (*EF-1α*) was measured with RT-qPCR to verify the RNA quality and quantity of aphid extractions.

For the siRNA–nanoparticle emulsion aerosolization trials, gene expression values were calculated using the standard curve method for relative quantification (Bookout *et al*., 2006) and normalized to the housekeeping gene *EF-1α* (Table S1) for *Ac. pisum* (Dunbar *et al*., 2007) and *S. graminum*. *Ap. glycines* gene expression was normalized to the housekeeping gene *ribosomal protein (RPS9)* (Bansal *et al*., 2012; Table S1).

Statistical analysis

For the GFP dsRNA degradation trials ANOVA was used to determine if there was a statistically significant difference in normalized expression between treatments (Bookout *et al*., 2006). Post-hoc multiple comparison analyses of normalized expression values were conducted between treatments using least significant difference tests. The statistical program R (R Core Team, 2016, R version 3.3.1) was used for all statistical analyses and an α of 0.05 or less was chosen a priori as the significance threshold for treatment differences.

For the nanoparticle spray trials, statistical analysis was performed using unpaired, one-tailed Welch's *t*-tests comparing each experimental treatment (siRNA + nanoparticles and siRNA only) with the control treatment (control siRNA + nanoparticles). Note that trials in which we observed a significant gene knockdown at an α of 0.05 or less and a corresponding change in phenotype were repeated. For the experiments in which we tested the effect of aerosolization on aphid mortality in *S. graminum* and *Ap. glycines*, a two-proportion *z*-test was used to determine if there were differences in percent aphid mortality. A *z*-test was used because we were making comparisons using proportion data.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments and suggestions. We would also like to thank Mike Scott for his help with providing us with nanoparticles and advice. We thank Cari Vanderpool's lab for donating the GFP plasmid to our lab for this study. We would also like to thank the undergraduate Yihui Zhu and technician Andrew Jang who helped with technical aspects of this project and all members of the Hansen lab who helped in various aspects. We would like to thank members of the Hansen and Degnan labs for their feedback on the manuscript. We would also like to thank Andy Michel for assistance in developing the Insect Genetic Technologies Research Coordination Network (IGTRCN) proposal and in various other aspects. We thank Carla Cáceres and members of her lab for use of the micro-scale. This work was supported by University of Illinois at Urbana-Champaign (UIUC) start-up funds to A.K.H., the project Postdoc USB (reg. no. CZ.1.07/2.3.00/30.0006) realized through the EU Education for Competitiveness Operational Programme to E.N., an IGTRCN grant to R.B. for peer-to-peer training in A.K.H.'s laboratory, The School of Integrative Biology at UIUC for the Robert H. Davis Undergraduate Research Prize to T.E.H., iBio summer internship to I.H.S., and Francis M. and Harlie M. Clark Summer Fellowship to M.W.T.

References

Bansal, R. and Michel, A. (2013) Core RNAi machinery and Sid1, a component for systemic RNAi, in the hemipteran insect, *Aphis glycines*. *Int J Mol Sci* **14**: 3786–3801.

Bansal, R., Mamidala, P., Mian, M.A.R., Mittapalli, O. and Michel, A.P. (2012) Validation of reference genes for gene expression

studies in *Aphis glycines* (Hemiptera: Aphididae). *J Econ Entomol* **105**: 1432–1438.

Bansal, R., Mian, M.A.R., Mittapalli, O. and Michel, A.P. (2014) RNA-Seq reveals a xenobiotic stress response in the soybean aphid, *Aphis glycines*, when fed aphid-resistant soybean. *BMC Genomics* **15**: 972.

Bellés, X. (2010) Beyond Drosophila: RNAi *in vivo* and functional genomics in insects. *Annu Rev Entomol* **55**(1): 111–128.

Blackman, R. L. and Eastop, V. F. (2006) *Aphids on the World's Herbaceous Plants and Shrubs*. Wiley, Chichester.

Bookout, A.L., Cummins, C.L., Mangelsdorf, D.J., Pesola, J.M. and Kramer, M.F. (2006) High-throughput real-time quantitative reverse transcription PCR. *Curr Protoc Mol Biol* Chapter 15, Unit 15.8.

Christiaens, O., Swevers, L. and Smagghe, G. (2014) DsRNA degradation in the pea aphid (*Acyrthosiphon pisum*) associated with lack of response in RNAi feeding and injection assay. *Peptides* **53**: 307–314.

Coleman, A.D., Wouters, R.H.M., Mugford, S.T. and Hogenhout, S.A. (2015) Persistence and transgenerational effect of plant-mediated RNAi in aphids. *J Exp Bot* **66**: 541–548.

Cooper, S.G., Concibido, V., Estes, R., Hunt, D., Jiang, G.-L., Krupke, C. *et al*. (2015) Geographic distribution of soybean aphid biotypes in the United States and Canada during 2008–2010. *Crop Sci* **55**: 2598.

Das, S., Debnath, N., Cui, Y., Unrine, J. and Palli, S.R. (2015) Chitosan, carbon quantum dot, and silica nanoparticle mediated dsRNA delivery for gene silencing in *Aedes aegypti*: a comparative analysis. *ACS Appl Mater Interfaces* **7**: 19530–19535.

Dunbar, H.E., Wilson, A.C.C., Ferguson, N.R. and Moran, N.A. (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* **5**: e96.

Guo, H., Song, X., Wang, G., Yang, K., Wang, Y., Niu, L. *et al*. (2014) Plant-generated artificial small RNAs mediated aphid resistance. *PLoS One* **9**: e97410.

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

Hansen, A.K., and Moran, N.A. (2012). Altered tRNA characteristics and 3' maturation in bacterial symbionts with reduced genomes. *Nucleic Acids Res*., **40**:7870–7884.

He, C., Hu, Y., Yin, L., Tang, C. and Yin, C. (2010) Effects of particle size and surface charge on cellular uptake and biodistribution of polymeric nanoparticles. *Biomaterials* **31**: 3657–3666.

Jaubert-Possamai, S., Le Trionnaire, G., Bonhomme, J., Christophides, G.K., Rispe, C. and Tagu, D. (2007) Gene knockdown by RNAi in the pea aphid *Acyrthosiphon pisum*. *BMC Biotechnol* **7**: 63.

Jaubert-Possamai, S., Rispe, C., Tanguy, S., Gordon, K., Walsh, T., Edwards, O. and Tagu, D. (2010) Expansion of the miRNA pathway in the hemipteran insect *Acyrthosiphon pisum*. *Mol Biol Evol* **27**: 979–987.

Kaneda, M.M., Sasaki, Y., Lanza, G.M., Milbrandt, J. and Wickline, S.A. (2010) Mechanisms of nucleotide trafficking during siRNA delivery to endothelial cells using perfluorocarbon nanoemulsions. *Biomaterials* **31**: 3079–3086.

Kim, K.S., Hill, C.B., Hartman, G.L., Rouf Mianc, M.A. and Diers, B.W. (2008) Discovery of soybean aphid biotypes. *Crop Sci* **48**: 923.

Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T.C., Kaneda, M. *et al*. (2013) RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc Natl Acad Sci USA* **110**: 12750–12755.

Mao, J. and Zeng, F. (2012) Feeding-based RNA interference of a gap gene is lethal to the pea aphid, *Acyrthosiphon pisum*. *PLoS ONE* **7**: e48718.

McCornack, B.P., Ragsdale, D.W. and Venette, R.C. (2004) Demography of soybean aphid (Homoptera: Aphididae) at summer temperatures. *J Econ Entomol* **97**: 854–861.

Moran, N.A. and Jarvik, T. (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624–627.

Mutti, N.S., Park, Y., Reese, J.C. and Reeck, G.R. (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. *J Insect Sci* **6**: 1–7.

NCBI (2017) National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/genome/browse/. Accessed on 12 January 2017.

Nováková, E. and Moran, N.A. (2012) Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Mol Biol Evol* **29**: 313–323.

Ortiz-Rivas, B., Jaubert-Possamai, S., Tanguy, S., Gauthier, J.-P., Tagu, D. and Claude, R. (2012) Evolutionary study of duplications of the miRNA machinery in aphids associated with striking rate acceleration and changes in expression profiles. *BMC Evol Biol* **12**: 216.

Pitino, M., Coleman, A.D., Maffei, M.E., Ridout, C.J. and Hogenhout, S.A. (2011) Silencing of aphid genes by dsRNA feeding from plants. *PLoS ONE* **6**: e25709.

Poliakov, A., Russell, C.W., Ponnala, L., Hoops, H.J., Sun, Q., Douglas, A.E. and van Wijk, K.J. (2011) Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. *Mol Cell Proteomics* **10**: M110.007039.

R Core Team (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, (R version 3.3.1) Vienna, Austria. https://www.R-project.org/.

Ross, K.A., Brenza, T.M., Binnebose, A.M., Phanse, Y., Kanthasamy, A.G., Gendelman, H.E. *et al*. (2015) Nano-enabled delivery of diverse payloads across complex biological barriers. *J Control Release* **219**: 548–559.

Scott, J.G., Michel, K., Bartholomay, L.C., Siegfried, B.D., Hunter, W.B., Smagghe, G. *et al*. (2013) Towards the elements of successful insect RNAi. *J Insect Physiol* **59**: 1212–1221.

Shabalina, S.A. and Koonin, E.V. (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* **23**: 578–587.

Shakesby, A.J., Wallace, I.S., Isaacs, H.V., Pritchard, J., Roberts, D.M. and Douglas, A.E. (2009) A water-specific aquaporin involved in aphid osmoregulation. *Insect Biochem Mol Biol* **39**: 1–10.

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

Wilson, A.C.C., Ashton, P.D., Calevro, F., Charles, H., Colella, S., Febvay, G. *et al*. (2010) Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol* **19**: 249–258.

Wynant, N., Santos, D. and Vanden Broeck, J. (2014) Biological mechanisms determining the success of RNA interference in insects. *Int Rev Cell Mol Biol* **312**: 139–167.

Yahia, E.M. and Ornelas-Paz, J.J. (2009) Chemistry, Stability, and Biological Actions of Carotenoids. *Fruit and Vegetable*

Phytochemicals, (de la Rosa, L., Alvarez-Parrilla, E., and González-Aguilar. G.A., eds.) pp. 177–222. Wiley-Blackwell, Oxford, UK. doi:10.1002/9780813809397.ch7.

Yu, N., Christiaens, O., Liu, J., Niu, J., Cappelle, K., Caccia, S. *et al.* (2013) Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect Sci* **20**: 4–14.

Zhang, X., Zhang, J. and Zhu, K.Y. (2010) Chitosan/double-stranded RNA nanoparticle-mediated RNA interference to silence chitin synthase genes through larval feeding in the African malaria mosquito (*Anopheles gambiae*). *Insect Mol Biol* **19**: 683–693.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Primers used in RNA interference and reverse transcriptase-quantitative PCR experiments.

Table S2. Aerosolized double-stranded green fluorescent protein RNA degradation trials and time points sampled.

Table S3. One-tailed *t*-tests of carotene dehydrogenase-short interfering RNA trials.

Table S4. One-tailed test of branched-chain amino acid transaminase (bcat) 100 nM trials*.

Table S5. One-tailed *t*-tests of branched-chain amino acid transaminase (bcat) 200 nM trials*.

Figure S1. Induction of the RNA interference silencing pathway via the introduction of short interfering RNA or double-stranded RNA.

Figure S2. Determination of the effects of an aerosolized water solution with and without nanoparticles (NPs) on aphid mass and overall mortality. Error bars indicate ± SE from the mean. Bars with different letters above represent significant differences within a trial at $P < 0.05$. For the mass data, a *t*-test was run between the two treatments (*Schizaphis graminum:* *t*-value = 0.97, $P = 0.17$; *Aphis glycines:* *t*-value = 0.44, $P = 0.67$). For the aphid mortality data a *z*-test was performed between the Water + NP treatment and Control (no spray) treatment (*S. graminum:* $z = 1.06$, $P = 0.1442$; *Ap. glycines:* *z*-score = 0.21, $P = 0.41$). The control treatment was used to determine background aphid mortality. There was no significant differences found between the Water + NP treatment and the Water-only treatment (*S. graminum:* *z*-score = 0.40, $P = 0.35$; *Ap. glycines:* *z*-score = 0.64, $P = 0.26$).

Figure S3. *Aphis glycines* aphids exposed to 200 nM *branched-chain amino acid transaminase* (*bcat*) short interfering RNA + 200 pM nanoparticles at 5 days old and collected 3 days after aerosolization treatment. Error bars indicate ± SE from the mean (n = 6 for each treatment). Bars with different letters above represent significant differences within a trial at $P < 0.05$. Relative gene expression was calculated using the comparative cycle threshold (Ct) method. Relative *bcat* gene expression was compared to expression of the housekeeping gene ribosomal protein *(RPS9)*.

Paper 4:

Nováková, E., Hypša, V., Nguyen, P., Husník, F., & Darby, A. C. (2016). Genome sequence of *Candidatus* Arsenophonus lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *Standards in genomic sciences*, *11*(1), 72.

Nováková et al. Standards in Genomic Sciences (2016) 11:72
DOI 10.1186/s40793-016-0195-1

Standards in Genomic Sciences

SHORT GENOME REPORT Open Access

Genome sequence of *Candidatus* Arsenophonus lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae)

Eva Nováková[1*], Václav Hypša[1], Petr Nguyen[2], Filip Husník[1] and Alistair C. Darby[3]

Abstract

Candidatus Arsenophonus lipopteni (*Enterobacteriaceae*, *Gammaproteobacteria*) is an obligate intracellular symbiont of the blood feeding deer ked, *Lipoptena cervi* (Diptera: Hippoboscidae). The bacteria reside in specialized cells derived from host gut epithelia (bacteriocytes) forming a compact symbiotic organ (bacteriome). Compared to the closely related complex symbiotic system in the sheep ked, involving four bacterial species, *Lipoptena cervi* appears to maintain its symbiosis exclusively with *Ca. Arsenophonus lipopteni*. The genome of 836,724 bp and 24.8 % GC content codes for 667 predicted functional genes and bears the common characteristics of sequence economization coupled with obligate host-dependent lifestyle, e.g. reduced number of RNA genes along with the rRNA operon split, and strongly reduced metabolic capacity. Particularly, biosynthetic capacity for B vitamins possibly supplementing the host diet is highly compromised in *Ca. Arsenophonus lipopteni*. The gene sets are complete only for riboflavin (B2), pyridoxine (B6) and biotin (B7) implying the content of some B vitamins, e.g. thiamin, in the deer blood might be sufficient for the insect metabolic needs. The phylogenetic position within the spectrum of known *Arsenophonus* genomes and fundamental genomic features of *Ca. Arsenophonus lipopteni* indicate the obligate character of this symbiosis and its independent origin within Hippoboscidae.

Keywords: *Arsenophonus*, Symbiosis, Tsetse, Hippoboscidae

Introduction

Symbiosis has for long been recognized as one of the crucial drivers of evolution. In insects, numerous symbiotic relationships, mainly with bacteria, enabled the hosts to exploit various environments and/or life strategies, and supposedly started adaptive radiations in some groups. The mechanisms of such evolutionary processes include for example contribution to the host immunity, modification of the reproductive strategy, or provision of essential compounds to the hosts relying on nutritionally compromised resources. Blood feeding (hematophagous) insects provide an illuminating example of a life strategy shift coupled with symbiosis. Since blood meal lacks some of the B vitamins, hematophagous insects rely on their supply by symbiotic bacteria. The relationships between bacteria and hematophagous insect displays considerable degree of variability spanning from less intimate associations with entire gut microbial community, e.g. triatomine bugs [1, 2], to highly specialized interactions with few or single obligate symbiont(s), e.g. lice, bed bugs, tsetse flies, louse flies and bat flies [3–7]. With the recent advancement of genomic approaches and genetic manipulations, symbioses in these insect groups, often important disease vectors, have become of a high interest.

Here we describe fundamental biological characteristics and genome properties of the obligate symbiont of a deer ked, *Lipoptena cervi* (Hippoboscidae). In comparison to multipartite symbiotic systems of closely related hosts from families Hippoboscidae (i.e. *Melophagus ovinus* [7]) and Glossinidae (i.e. *Glossina* sp. [6]), *Lipoptena*

* Correspondence: novaeva@paru.cas.cz
[1]Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre, ASCR, v.v.i. Branisovska 31, 37005 Ceske Budejovice, Czech Republic
Full list of author information is available at the end of the article

cervi harbors a single unaccompanied obligate symbiont from the genus *Arsenophonus*. The genome of *Candidatus* Arsenophonus lipopteni has been sequenced for two reasons. The first was to extend our knowledge on occurrence and genomics of the obligate symbionts across the spectrum of hematophagous hosts involved in strictly bilateral symbiosis, e.g. bed bugs [5], head lice [4], leaches [8, 9]. This is a necessary prerequisite for the future analysis of the origins and evolution of this kind of symbioses. In addition, we intend to use the sequence in a broader comparative framework focused on evolution of bacterial symbiosis, particularly on its role in B vitamin provision to various ecological types of the hosts.

Organism information

Classification and features

Ca. Arsenophonus lipopteni has an obligate association with its host, *L. cervi*, and is therefore uncultivable. In order to localize the bacteria within the host, Fluorescent In Situ Hybridization and Transmission Electron Microscopy was performed on dissected gut tissue as described in detail in [7]. For FISH, the tissue was fixed and hybridized in tubes with eubacterial (EUB338, Flc-GCTGCCTCCCGTAGGA; [10]) and *Ca. Arsenophonus lipopteni* specific probes (ArL, Cy3-CTGACTAACGC TTGCACC; this study). The later was designed in a variable region of 16S rRNA gene taking the target sequence accessibility into account [11].

The distribution of *Ca. Arsenophonus lipopteni* (Fig. 1) in the host body closely resembles that of *Ca. Arsenophonus* melophagi and *Wigglesworthia glossinidia*^{T}, the obligate symbionts of the blood sucking flies *Melophagus ovinus* and *Glossina* sp., respectively [7]. Highly pleomorphic cells of the Gram negative non-sporulating bacteria from the family *Enterobacteriales* are primarily found in the modified part of the gut wall (bacteriome) formed by the specialized enlarged epithelial cells (bacteriocytes, Fig. 1c, 1d). Additional key features of *Ca. Arsenophonus lipopteni* are provided as a standardized summary in Table 1.

Apart from the functional characterization, the genome sequence of Ca. *Arsenophonus lipopteni* was also utilized to assess the relationship of this bacterium to other *Arsenophonus* symbionts. Since the sequence compositional shift compromises phylogenetic usage of 16S rDNA, leading to topological artifacts with long branched symbiotic taxa clustering together [12], we carried out a phylogenetic analysis of a multi-gene matrix and used advanced Bayesian approaches. The matrix was generated for all available *Arsenophonus* genomes (incl. *Ca. Riesia pediculicola*), five other symbionts, eight non-symbiotic members of *Enterobacteriaceae*, and two outgroups. A set of 70 orthologous genes was determined as an intersection of COGs shared by these bacteria (generated using the MicrobesOnline database; [13]) with "SICO" gene list [14]. The genes were retrieved from the finished assembly using Blastp searches [15] and processed as described previously [7]. The resulting matrix contained 22618 unequivocally aligned positions. Phylobayes [16], a tool specifically developed to overcome the difficulty with heterogeneous composition of sequences, was used for the tree reconstruction. The analysis was run in 2 chains under the GTR + CAT model with amino acids recoded according to the Dayhoff6 option. When the convergence was not reached after 20,000 cycles, the program was stopped and majority rule consensus was calculated after discarding 4,000 cycles burn-in.

The results confirm *Ca. Arsenophonus lipopteni* membership in the genus *Arsenophonus*. All *Arsenophonus* species (including *Ca.* Rieisa pediculicola) formed a well-supported monophyletic branch clustering as a

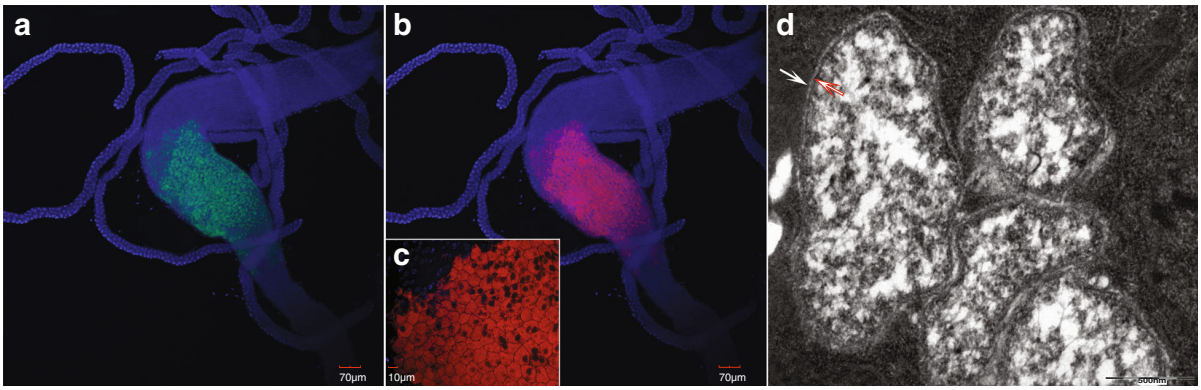


Fig. 1 Visualization of the bacteria in the host tissue using FISH and TEM (D). The symbiotic organ (bacteriome) localized in the midgut section harboring *Ca. Arsenophonus lipopteni* targeted with green (Flc) labeled eubacterial probe (**a**) and red (Cy3) labeled specific probe (**b**). Detail of the host cells (bacteriocytes) filled with the symbionts (**c**). The blue signal is DAPI stained DNA. Four cells of *Ca. Arsenophonus lipopteni* under TEM (**d**). The white arrow points to bacterial outer membrane and the red bordered arrow shows the cytoplasmatic cell membrane

Table 1 Classification and general features of *Ca.* Arsenophpnus lipopteni

MIGS ID	Property	Term	Evidence code[a]	
	Classification	Domain *Bacteria*	TAS	[33]
		Phylum *Proteobacteria*	TAS	[34]
		Class *Gammaproteobacteria*	TAS	[35]
		Order "*Enterobacteriales*"	TAS	[36]
		Family *Enterobacteriaceae*	TAS	[37]
		Genus *Arsenophonus*	TAS	[38]
		Species *Ca. Arsenophonus lipopteni*	IDA	
		Strain: CB	IDA	
	Gram stain	Negative	TAS	[38]
	Cell shape	Pleomorphic	NAS	
	Motility	Non-motile	TAS	[38]
	Sporulation	Non-sporulating	TAS	[38]
	Temperature range	Not determined	IDA	
	Optimum temperature	Not determined	IDA	
	pH range; Optimum	Not determined	IDA	
	Carbon source	Not determined	IDA	
MIGS-6	Habitat	Insect host; bacteriome of *L. cervi*	IDA	
MIGS-6.3	Salinity	Not determined	IDA	
MIGS-22	Oxygen requirement	Not determined	IDA	
MIGS-15	Biotic relationship	Symbiotic	IDA	
MIGS-14	Pathogenicity	Non-pathogen	NAS	
MIGS-4	Geographic location	Ceske Budejovice, Czech Republic	IDA	
MIGS-5	Sample collection date	June 2010	IDA	
MIGS-4.1	Longitude	14.43	IDA	
MIGS-4.2	Latitude	48.97	IDA	
MIGS-4.4	Altitude	399 m	IDA	

[a]Evidence codes, IDA Inferred from Direct Assay, TAS Traceable Author Statement (i.e., a direct report exists in the literature), NAS Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [39]

sister group to *Providencia* (Fig. 2). Despite the length of branches for the obligate symbionts with highly modified genomes, this arrangement was assigned a high posterior probability. Although the six included *Arsenophonus* lineages certainly do not form a monophyletic group within the known *Arsenophonus* spectrum [17], the results indicate that *Ca. Arsenophonus lipopteni* evolved independently from *Ca. Arsenophonus* melophagi housed in related Hippoboscidae host [7].

Genome sequencing information

Genome project history

The host specimens *Lipoptena cervi* were collected from wild populations during summer 2010 in the Czech Republic. Finished genome sequence has been deposited in GenBank under acc. No. CP013920 on January 11, 2016. A summary on the sequencing project is provided in Table 2.

Growth conditions and genomic DNA preparation

Since the bacterium is uncultivable, the host tissue was used for DNA extraction. The gut tissue containing the symbiotic organs were dissected from 6 flies in 1× phosphate buffered saline, homogenized with a sterile mortar and pestle and extracted using QiaAmp DNA Micro Kit (QIAGEN, United Kingdom). The DNA quality was assessed using Agilent 2100 Bioanalyzer (Agilent Technologies).

Genome sequencing and assembly

The paired end 100 bp long reads were generated on one lane of Illumina HiSeq2000 run at Yale Center for Genome Analysis. A5 assembly pipeline with the default settings was used to assemble the reads [18]. Of the 109,640 resulting contigs, the longest contig (836,730 bp) with 40× fold coverage formed a circular molecule with 99 bp overlap at the ends. This contig corresponds to the *Ca.*

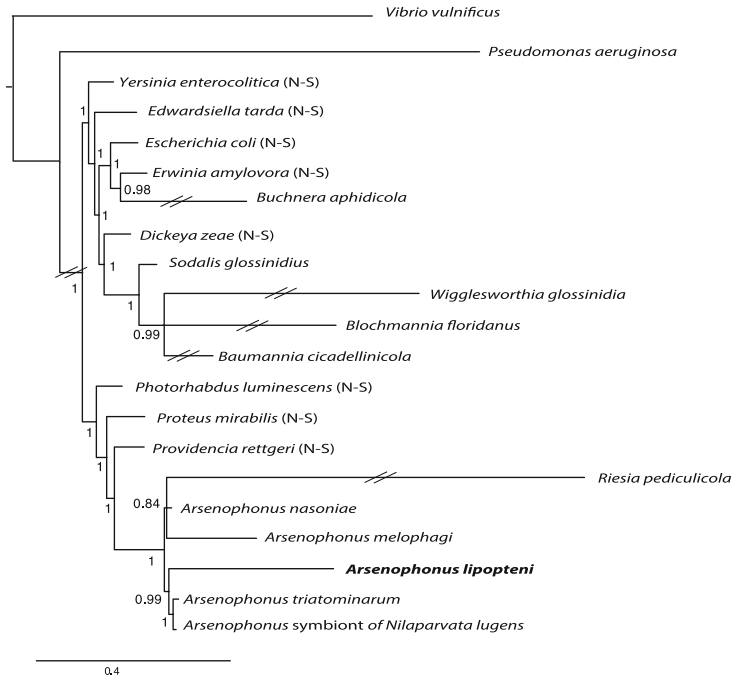


Fig. 2 Phylogenomic reconstruction of *Ca. Arsenophonus lipopteni* position. The length of the double crossed branches was scaled to 1:4. The numbers indicate posterior probability for each node. The "N-S" in the brackets following the taxon name designates the non-symbiotic bacteria included into the dataset

Arsenophonus lipopteni genome. Pilon v1.12 [19] was used to check assembly quality and to improve base calls and small indels.

Genome annotation

The finished genome was annotated using a combination of following tools: RAST [20], PGAP, and Prokka v1.10 [21]. The annotation was then manually curated and checked for the presence of gene remnants. The final annotation is available in GenBank (CP013920). Metabolic pathways were reconstructed in the RAST server [20] and gene absence was verified using BlastP searches. Proteins were assigned to the clusters of orthologous groups using COGnitor [22], and the presence of signal peptides was detected using SignalP [23]. Pfam domains were predicted using HMMER [24] against the Pfam-A database [25]. Transmembrane predictions were done using TMHMM Server v. 2.0. The search for CRISPR repeats was performed in Geneious [26].

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	2 × 100 bp paired end
MIGS 29	Sequencing platforms	Illumina
MIGS 31.2	Fold coverage	40
MIGS 30	Assemblers	A5
MIGS 32	Gene calling method	RAST, PGAP, PROKKA
	Locus Tag	AUT07
	Genbank ID	CP013920
	Genbank Date of Release	1-25-2016
	GOLD ID	Gp0127464
	BIOPROJECT	PRJNA306001
MIGS 13	Source Material Identifier	Host tissue
	Project relevance	Evolution of bacterial symbiosis

Genome properties

The finished genome consists of 836,724 nucleotides in a single circular chomosome with a low GC content of 24.9 %. The total number of predicted functional genes (667) relative to the genome size implies a lower coding density (75.8 %). The average gene length of 1,001 bp however does not suggest that the genome underwent an extreme economization typical for the obligate symbionts, e.g. *Buchnera aphidicola* str. Cc, *Ca. Sulcia muelleri*, *Ca. Carsonella ruddii*, *Ca. Zinderia insecticolla* [27]. Over 99 % of protein coding genes have been assigned to particular COGs and Pfam domains (Tables 3 and 4). Signal peptides and transmembrane helices have been identified for 5 and 124 protein coding genes respectively (Table 3). The noncoding RNA genes consist

Table 3 Statistics for finished genome assebly of *Ca. Arsenophonus lipopteni*

Attribute	Value	% of Total[a]
Genome size (bp)	836,724	100.00
DNA coding (bp)	633,822	75.80
DNA G + C (bp)	208,103	24.90
DNA scaffolds	1	100.00
Total genes	683	100.00
Protein coding genes	625	91.50
RNA genes	42	6.10
Pseudo genes	16	2.30
Genes assigned to COGs	622	99.52
Genes assigned Pfam domains	625	100.00
Genes with signal peptides	5	0.80
Genes with transmembrane helices	124	19.8.00
CRISPR repeats	0	0.00

[a]The total is based on either the size of the genome in bp or the total number of genes

Table 4 Number of protein coding genes assigned to the COG categories

Cat. code	Value	Percentage of total	Description
J	128	20.48	Translation, ribosomal structure and biogenesis
A	1	0.16	Processing and modification
K	17	2.72	Transcription
L	40	6.4	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	15	2.4	Cell cycle control, cell division, chromosome partitioning
Y	0	0	Nuclear structure
V	7	1.12	Defense mechanisms
T	10	1.6	Signal transduction mechanisms
M	70	11.2	Cell wall/membrane biogenesis
N	1	0.16	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	12	1.92	Intracellular trafficking and secretion
O	42	6.72	Posttranslational modification, protein turnover, chaperones
C	22	3.52	Energy production and conversion
G	26	4.16	Carbohydrate transport and metabolism
E	34	5.44	Aminoacid transport and metabolism
F	24	3.84	Nucleotide transport and metabolism
H	49	7.84	Coenzyme transport and metabolism
I	26	4.16	Lipid transport and metabolism
P	24	3.84	Inorganic ion transport and metabolism
Q	1	0.16	Secondary metabolites biosynthesis, transport and catabolism
R	9	1.44	General function prediction only
S	13	2.08	Function unknown
-	51	8.16	Assigned to more than one category
-	3	0.48	Not in COGs

of tmRNA, RNAseP, Alpha RBS, cspA, 35 tRNAs, and 3 rRNA genes (altogether 42 RNA genes). The three ribosomal genes are however not organized into a single operon, a phenomenon previously described for at least 9 unrelated bacterial clades, including gammaproteobacterial symbionts of the genus *Buchnera* and *Candidatus* Blochmannia, and attributed to their host-dependent lifestyle [28].

The genome properties described above coupled with 16 pseudogenes identified in the genome suggest rather recent establishment of the obligate symbiosis resulting in significant but recent gene/function loss without removal of presently non-coding regions. Regarding the coding capacity for B vitamins and related cofactors, the genome of *Ca. Arsenophonus lipopteni* appears to be highly economized. Similar to *Ca. Arsenophonus* melophagi, the bacteria cannot synthesize thiamine (B1), niacin (B3), panthothenic acid (B5) and folic acid (B9). In addition, the genome does not code for heme biosynthesis. Other basic genome characteristics are summarized in Table 3.

Conclusions

Compared to the closely related complex symbiotic system in the sheep ked, *Melophagus ovinus*, *Lipoptena cervi* appears to maintain symbiosis exclusively with *Ca. Arsenophonus lipopteni*. The growing number of genome sequences available for the symbionts and the hematophagous hosts involved in strictly bilateral symbiosis (e.g. [29, 30]) will help elucidating some common requirements on B vitamins, or possibly highlight diverse needs of insects digesting blood of various vertebrates. *Ca. Arsenophonus lipopteni* possesses complete gene sets for biosynthesis of three B vitamins, riboflavin (B2), pyridoxine (B6) and biotin (B7). While the metabolic capacity is directly assessed from genomic data, the presence of any vitamin efflux systems cannot be easily elucidated due to yet poorly understood mechanisms for vitamin export [31]. However, based on recent findings from other hematophagous systems, it has become more clear that the nutritional interaction does not rely on biosynthesis of all B vitamins as originally suggested by Puchta [32]. For instance, similar to all the other *Arsenophonus* genomes, biosynthetic capacity for thiamin is compromised in *Ca. Arsenophonus lipopteni*. The genome however possesses ABC thiamin transporter genes (*thiP, thiQ, tbpA*) implying the content of thiamin or thiamin pyrophosphate, compared to

e.g. biotin or riboflavin, in the host blood might be sufficient for the insect metabolic needs (Novakova, unpublished data). Within the spectrum of known *Arsenophonus* genomes ranging from 0.57 Mb of *Ca. Riesia pediculicola* to 3.5 Mb of *A. nasoniae*, representing various symbiotic types, the genomic sequence of *Ca. Arsenophonus lipopteni* clearly reflects characteristics common for obligate mutualists. Furthermore, the phylogenetic reconstruction suggests an independent origin of this obligate association within Hippoboscidae.

Acknowledgements
This work was supported by grant 13-01878S (Czech Science Foundation) to VH, and the project Postdoc USB (reg.no.CZ.1.07/2.3.00/30.0006) realized through EU Education for Competitiveness Operational Programme. PN was supported by Grant 14-35819P of the Czech Science Foundation.

Authors' contributions
EN, AD, VH participated in the design of the study and coordination. EN, FH and PN performed the imaging. VH and FH assembled and annotated the genome sequence. All authors participated in editing of the manuscript and read and approved the final version.

Competing interests
The authors declare that they have no involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

Author details
[1]Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre, ASCR, v.v.i. Branisovska 31, 37005 Ceske Budejovice, Czech Republic. [2]Faculty of Science, University of South Bohemia, and Institute of Entomology, Biology Centre, ASCR, v.v.i. Branisovska 31, 37005 Ceske Budejovice, Czech Republic. [3]Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK.

Received: 26 January 2016 Accepted: 7 September 2016
Published online: 17 September 2016

References

1. Beard CB, Cordon-Rosales C, Durvasula RV. Bacterial Symbionts of the Triatominae and their potential use in control of Chagas disease transmission 1. Annu Rev Entomol. 2002;47:123–41.
2. Gumiel M, da Mota FF, Rizzo Vde S, Sarquis O, de Castro DP, Lima MM, et al. Characterization of the microbiota in the guts of *Triatoma brasiliensis* and *Triatoma pseudomaculata* infected by *Trypanosoma cruzi* in natural conditions using culture independent methods. Parasit Vectors. 2015;8:245.
3. Hypša V, Křížek J. Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). Microbial Ecol. 2007;54:242–51.
4. Sasaki-Fukatsu K, Koga R, Nikoh N, Yoshizawa K, Kasai S, Mihara M, et al. Symbiotic bacteria associated with stomach discs of human lice. Appl Environ Microbiol. 2006;72:7349–52.
5. Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A. 2010;107:769–74.
6. Aksoy S, Chen X, Hypsa V. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). Insect Mol Biol. 1997;6:183–90.
7. Nováková E, Husník F, Šochová E, Hypša V. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. Appl Environ Microbiol. 2015;81:6189–99.
8. Nelson M, Graf J. Bacterial symbioses of the medicinal leech *Hirudo verbana*. Gut Microbes. 2012;3:322–31.
9. Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech *Haementeria officinalis*. Gen Biol Evol. 2015;7:2871–84.
10. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol. 1990;56:1919–25.
11. Behrens S, Rühland C, Inácio J, Huber H, Fonseca A, Spencer-Martins I, et al. In situ accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes. App Environ Microbiol. 2003;69:1748–58.
12. Charles H, Heddi A, Rahbe Y. A putative insect intracellular endosymbiont stem clade, within the *Enterobacteriaceae*, inferred from phylogenetic analysis based on a heterogeneous model of DNA evolution. C R Acad Sci Ser III Sci Vie. 2001;324:489–94.
13. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 2010;38:D396–400.
14. Lerat E, Daubin V, Moran NA. From gene trees to organismal phylogeny in prokaryotes: the case of the γ-proteobacteria. PLoS Biol. 2003;1:e19.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
16. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 2009; 25:2286–8.
17. Nováková E, Hypša V, Moran NA. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol. 2009;9:1–14.
18. Tritt A, Eisen JA, Facciotti MT, Darling AE. An integrated pipeline for de novo assembly of microbial genomes. PLoS One. 2012;7:e42304.
19. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9:e112963.
20. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
21. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
22. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.
23. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
24. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195.
25. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The Pfam protein families database. Nucleic Acids Res. 2002;30:276–80.
26. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–9.
27. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 2012;10:13–26.
28. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct. 2009;4:13.
29. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A. 2010;107:12168–73.
30. Attardo GM, Abila PP, Auma JE, Baumann AA, Benoit JB, Brelsfoard CL, et al. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. Science. 2014;344:380–6.
31. Jaehme M, Slotboom DJ. Diversity of membrane transport proteins for vitamins in bacteria and archaea. Biochimica et Biophysica Acta (BBA)-General Subjects. 2015;1850:565–76.
32. Puchta O. Experimentelle Untersuchungen über die Bedeutung der Symbiose der Kleiderlaus Pediculus vestimenti Burm. Z Parasitenkd. 1955;17:1–40.
33. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.
34. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, vol. Volume 2, Part B. 2nd ed. New York: Springer; 2005. p. 1.

35. Garrity GM, Bell JA, Lilburn T. Class III. *Gammaproteobacteria* class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, vol. Volume 2, Part B. 2nd ed. New York: Springer; 2005. p. 1.
36. Garrity GM, Holt JG. Taxonomic Outline of the Archaea and Bacteria. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's Manual of Systematic Bacteriology, vol. Volume 1. 2nd ed. New York: Springer; 2001. p. 155–66.
37. Judicial Commission Conservation of the family name *Enterobacteriaceae*, of the name of the type genus, and designation of the type species OPINION NO. 15. Int Bull Bacteriol Nomencl Taxon. 1958;8:73–4.
38. Werren JH. *Arsenophonus*. Bergey's Manual of Systematic Bacteriology (Vol. 2), G.M. Garrity (ed), New York: Springer-Verlag; 2004:1–4
39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.

Paper 5:

Nováková, E., Husník, F., Šochová, E., & Hypša, V. (2015). *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and environmental microbiology*, *81*(18), 6189-6199.

AEM
Journals.ASM.org

Arsenophonus and *Sodalis* Symbionts in Louse Flies: an Analogy to the *Wigglesworthia* and *Sodalis* System in Tsetse Flies

Eva Nováková,[a,b] Filip Husník,[a,b] Eva Šochová,[a] Václav Hypša[a,b]

Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic[a]; Institute of Parasitology, Biology Centre, ASCR, v.v.i., Ceske Budejovice, Czech Republic[b]

Symbiosis between insects and bacteria result in a variety of arrangements, genomic modifications, and metabolic interconnections. Here, we present genomic, phylogenetic, and morphological characteristics of a symbiotic system associated with *Melophagus ovinus*, a member of the blood-feeding family Hippoboscidae. The system comprises four unrelated bacteria representing different stages in symbiosis evolution, from typical obligate mutualists inhabiting bacteriomes to freely associated commensals and parasites. Interestingly, the whole system provides a remarkable analogy to the association between *Glossina* and its symbiotic bacteria. In both, the symbiotic systems are composed of an obligate symbiont and two facultative intracellular associates, *Sodalis* and *Wolbachia*. In addition, extracellular *Bartonella* resides in the gut of *Melophagus*. However, the phylogenetic origins of the two obligate mutualist symbionts differ. In *Glossina*, the mutualistic *Wigglesworthia* appears to be a relatively isolated symbiotic lineage, whereas in *Melophagus*, the obligate symbiont originated within the widely distributed *Arsenophonus* cluster. Although phylogenetically distant, the two obligate symbionts display several remarkably similar traits (e.g., transmission via the host's "milk glands" or similar pattern of genome reduction). To obtain better insight into the biology and possible role of the *M. ovinus* obligate symbiont, "*Candidatus* Arsenophonus melophagi," we performed several comparisons of its gene content based on assignments of the Cluster of Orthologous Genes (COG). Using this criterion, we show that within a set of 44 primary and secondary symbionts, "*Ca.* Arsenophonus melophagi" is most similar to *Wigglesworthia*. On the other hand, these two bacteria also display interesting differences, such as absence of flagellar genes in *Arsenophonus* and their presence in *Wigglesworthia*. This finding implies that a flagellum is not essential for bacterial transmission via milk glands.

Evolution of insect-bacterium symbiosis has resulted in a variety of associations in a broad range of insect taxa. Many traits of these associations (e.g., specifics of the arrangement of the host and symbiont metabolism, location of the symbionts, etc.) have been shown to reflect the ecological type of the symbiosis (i.e., obligate mutualism versus facultative symbiosis) and the host's nutritional demands due to the insufficiency of the diet (1–3). For example, phytophagous insects feeding on plant sap (e.g., aphids, whiteflies, or psyllids) and exclusively hematophagous insects (e.g., tsetse flies, sucking lice, and bedbugs) are the two most frequently studied ecological groups harboring obligate mutualistic symbionts (1, 3). However, compared to the large amount and complexity of the data accumulated on the sap-feeding insects, our knowledge on the symbioses in hematophagous insects is still limited, despite the number of species playing a crucial role as vectors for numerous pathogens, often causing a major burden to public health and world economies.

Blood feeding originated independently in several insects groups. Particularly, flies (Diptera) accommodate a large number of blood-feeding specialists. Among them, four families of the Hippoboscoidea group, i.e., Glossinidae (tsetse flies), Hippoboscidae (louse flies), Nycteribiidae, and Streblidae (bat flies) form a monophyletic cluster (4) and share another fundamental biological feature, unique among insects, adenotrophic viviparity (5). Larvae of these flies undergo their development in the female's uterus and are deposited shortly before pupation. Their nutrition is provided by the mother via accessory glands (called milk glands), tubular organs opening into the uterus (6).

In tsetse flies (Glossinidae), the blood-feeding strategy together with the viviparous reproductive mode gave rise to a multipartite symbiotic system. The two principal symbionts are the obligate mutualist *Wigglesworthia glossinidia*, housed intracellularly within a bacteriome organ adjacent to the anterior midgut (7), and a facultative symbiont, *Sodalis glossinidius*, present in almost all tissues (8). Both symbionts are vertically transmitted via milk glands (7, 9, 10) and are presumed to compensate for the nutritionally unbalanced blood diet (11–15). This highly specific microbiome of tsetse flies is usually accompanied by reproductive manipulators from the genus *Wolbachia* (16) and a diversity of other transient bacteria of unknown relationship to the host (17, 18). In contrast to Glossinidae, the family Hippoboscidae is a species-rich, highly diversified, and cosmopolitan group feeding on mammals and birds (4). Individual genera and species differ in their host specificities and life styles. While some resemble tsetse flies in being temporal parasites that seek the host only for short-term blood feeding, others shed their wings after finding the host,

Received 6 May 2015 **Accepted** 23 June 2015
Accepted manuscript posted online 6 July 2015
Citation Nováková E, Husník F, Šochová E, Hypša V. 2015. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. Appl Environ Microbiol 81:6189–6199. doi:10.1128/AEM.01487-15.
Editor: H. Goodrich-Blair
Address correspondence to Eva Nováková, novaeva@paru.cas.cz.
E.N. and F.H. contributed equally to this article.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01487-15.

doi:10.1128/AEM.01487-15

and some retain only highly reduced wings or may even be permanently wingless.

During our previous research, we have shown that members of the family Hippoboscidae host several symbiotic bacteria (19–21). For instance we have identified *Sodalis* and *Arsenophonus* bacteria as common associates of Hippoboscidae, likely experiencing multiple origins throughout the host evolution (21). In the manuscript, we characterized the symbiotic system of the sheep ked, *Melophagus ovinus* (Hippoboscidae), a Palearctic wingless species that spends its entire life on the host and is transmitted only by physical contact between sheep (22). Living exclusively on vertebrate blood that lacks B vitamins, the sheep ked requires obligate symbiotic bacteria for its successful growth and/or reproduction. To provide insight into the overall arrangement of this symbiosis, we present results from transmission electron microscopy (TEM) and fluorescent *in situ* hybridization (FISH) and describe morphological properties of the symbiotic organ, the bacteriome, and other tissues harboring symbionts. Furthermore, using a shotgun metagenomic approach, we identify the principal symbionts, members of the genus *Arsenophonus* and *Sodalis*, and provide their phylogenetic and genomic characteristics. To fully understand interactions within the system, we also considered two additional bacteria present in the sheep keds, *Bartonella melophagi* (here referred to as *B. melophagi*) and *Wolbachia* sp. While we obtained genome assemblies for *Arsenophonus*, *Sodalis*, and *Bartonella* symbionts, *Wolbachia* raw data were of extremely low coverage (approximately 8-fold) and were thus not further analyzed.

Since we show that the symbiotic species "*Candidatus* Arsenophonus melophagi" and "*Candidatus* Sodalis melophagi" (here referred to as "*Ca*. Arsenophonus" and "*Ca*. Sodalis," respectively) have originated independently of the symbionts in tsetse flies, we suggest that their comparison to the symbionts in Glossinidae may help in distinguishing phylogenetic constraints, adaptive processes, and contingent traits. We also propose *M. ovinus* as a potential comparative model for further investigation into symbiosis of the blood-feeding insect groups with adenotrophic viviparity.

MATERIALS AND METHODS

Insect samples, dissections, light microscopy, and TEM. Samples of *M. ovinus* were obtained from a licensed family sheep farm at Krásetín, Czech Republic (48°53′4.549″N, 14°18′51.894″E). Bacteriomes and milk glands were dissected in phosphate-buffered saline (PBS) under an Olympus SZ51 dissecting microscope. The bacteriome tissue samples were used for both DNA extraction and fluorescent *in situ* hybridization (FISH). Transmission electron microscopy (TEM) of milk glands was carried out as reported previously for bacteriomes (19).

Sample preparation, probe design, and FISH. The tissue samples including gut and reproductive tract with surrounding milk glands were fixed overnight in 4% paraformaldehyde and then incubated in 2% H_2O_2 for up to 48 h to quench tissue autofluorescence. Afterwards, tissues were embedded in paraffin, and 6-μm-thick sections were placed on silane-coated slides. The slides were kept dry at a room temperature until used in fluorescent *in situ* hybridization (FISH). The probes were designed to match specifically the 16S rRNA of each associated bacteria to avoid any cross-hybridization. The labeled probe sequences are as follows: "*Ca*. Arsenophonus," Flc-GGCCTTACGGTCCCTCA (where Flc is fluorescein); "*Ca*. Sodalis," Cy3-CATCGCCTTCCTCCCCGCTG; *B. melophagi*, TxRd-CGTCATTATCTTCACCGGTG (where TxRd is Texas Red); *Wolbachia* sp., Cy5-CTTCTGTGAGTACCGTCATTATC and Cy3-AACGCTAGCC CTCTCCGTA; these allow various probe combinations in a single hybridization. The sections were dewaxed in xylene for 10 min, followed by 10 min in a xylene-ethanol equal-ratio solution. After rehydration in a graded ethanol series, the samples were treated with a drop of 70% acetic acid and incubated at 45°C until dry. The samples were then dehydrated through a graded ethanol series and underwent deproteinization with proteinase K solution (Qiagen) for 8 min at 45°C. The slides were rinsed in PBS, dehydrated, and air dried prior to the prehybridization step, which was carried out for 30 min at 45°C using hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl, pH 7.2, 0.1% SDS). Hybridization was performed for 3 h at 45°C directly on slides coated with 200 μl of hybridization buffer with 2 ng/μl of specific, labeled probe for each of the bacterial associates. Afterwards, the slides were rinsed twice for 5 min in the hybridization buffer and briefly rinsed in PBS and water. The slides were mounted in Vectashield medium (Vector Laboratories) containing 4′,6-diamidino-2-phenylindole (DAPI) and kept in the dark at 4°C until observed on an Olympus FV1000 confocal microscope. Negative controls (using RNase-treated slides) and controls for probe specificity (employing a general eubacterial probe EUB338, Flc-GCTGCCTCCCGTAGGA) (23) were performed along with each hybridization (data not shown).

DNA isolation and PCR. Total genomic DNA (gDNA) for construction of an Illumina paired-end (PE) library was obtained from a single bacteriome. DNA was extracted using a QIAamp DNA microkit (Qiagen) according to the standard kit protocol. DNA concentration was determined using Qubit Fluorometric Quantification (Life Technologies), and its quality was assessed in 1% agarose gel using standard electrophoresis. In total, 3 μg of gDNA was used for PE library construction.

DNA isolates acquired from whole individuals were examined for "*Ca*. Sodalis," *B. melophagi*, and *Wolbachia* presence with diagnostic PCR and further used for *Wolbachia* multilocus sequence typing (MLST). MLST genes from *Wolbachia* were sequenced according to a *Wolbachia* MLST protocol (24). Diagnostic PCR with specific primers was performed on 82 flies (55 males and 27 females). Details on primer sequences, either newly designed or retrieved from literature (25, 26), and amplicon lengths are presented in Table S5 in the supplemental material. In order to standardize the PCR conditions, ready-to-use PPT master mix (Top Bio) was employed, and the sensitivity of PCR was evaluated on five different DNA concentrations ranging from 5 to 60 ng/μl. Products of several diagnostic PCRs were Sanger sequenced to ensure primer specificity.

"*Ca*. Sodalis" gDNA for Illumina 3-kb mate-pair (MP) library construction was extracted from a single colony cultivation of "*Ca*. Sodalis" spread on Mitsuhashi and Maramorosch insect (MMI) medium blood plate (19) using the same procedure. In total, 10 μg of gDNA was used for MP library construction.

Genome assemblies. Bacteriome gDNA was sequenced using an Illumina PE library with an average insert size of 180 bp. The PE and MP libraries were sequenced bidirectionally on one lane of a HiSeq 2000 system on 100-bp and 75-bp runs, respectively (Keck Microarray Resource, West Haven, CT, USA). In total, over 80 million reads were obtained from the PE library. Reads underwent adaptor and quality trimming, and those shorter than 25 bp were discarded. Processed reads were assembled as metadata using two different approaches. First, *de novo* assembly of all the processed reads into contigs was achieved on a CLC Genomic Workbench (CLC bio A/S) with default parameter settings. Second, a subset of roughly 50 million reads consisting of paired reads only was randomly reduced in order to adjust the average coverage for *Arsenophonus* data to 100-fold and assembled using Velvet (27). *De novo* assembly in Velvet was performed with a *k*-mer size set to 55 bp. Contigs from each assembly were binned separately based on their average coverage and blastx (28) hits against a nonredundant (nr) database. Two sets of almost identical *Arsenophonus* binned data were then compared and combined into the draft genome of 40 *Arsenophonus* contigs. The contigs were submitted to an automatic annotation pipeline of the Joint Genome Institute, Integrated Microbial Genomes/Expert Review (IMG/ER) (29), using Prodigal (30) as the preferred gene calling method.

A draft genome of "*Ca*. Sodalis" was assembled from a combination of PE and MP Illumina data by the Velvet assembler (27), with all parameters

determined by VelvetOptimiser (http://www.vicbioinformatics.com/software.velvetoptimiser.shtml) and default scaffolding. Automatic annotation for comparative purposes was executed in RAST (http://rast.nmpdr.org). A *B. melophagi* draft genome was assembled from PE data by following the same procedure. A *k*-mer size of 71 was used for *B. melophagi*, and a *k*-mer size of 57 was used for "*Ca.* Sodalis" because a 75-bp size was used for MP reads. The "*Ca.* Sodalis" genome was further scaffolded by SSPACE (31). Gaps in *B. melophagi* scaffolds were closed by PCR. The completeness of gammaproteobacterial draft genomes was evaluated according to the number of single-copy (SICO) genes (32) present in the assemblies assessed by blastp and tblastn searches (28).

Genome comparisons. Initial genome comparisons between "*Ca.* Arsenophonus" and *Wigglesworthia glossinidia* and between "*Ca.* Sodalis" and *S. glossinidius* were based on the elementary genome characteristics, e.g., genome size, GC content, number of coding sequences (CDS), and tRNA genes. For the more detailed analysis, orthologous gene clusters were used as the comparative data. This comparison, performed in a broader phylogenetic spectrum, was based on gene numbers within functional ortholog assignments as defined by the Cluster of Orthologous Genes (COG) database and implemented in the web project IMG (29). The data set was composed of gene numbers for 42 bacterial genomes retrieved by the abundance profile function in IMG/ER (29) and supplemented with the data from the genomes of the *Wolbachia* endosymbiont of *Cimex lectularius*, *Wolbachia* wCle, and "*Ca.* Sodalis" missing in IMG/ER (see Table S1 in the supplemental material). For the two latter genomes, the COG assignment was performed using the COG database (ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd) and reverse position-specific (RPS)-BLAST, which allows for comparison of a query against a library of position-specific scoring matrices (PSSMs) (33). Altogether, 2,473 orthologous clusters were included in the comparison. The analysis was based on two different measures, one expressed as the number of the genes present in the genomes and the other as a percentage of these genes calculated from the sum of all COG clusters. A dissimilarity matrix was calculated from these data using Bray-Curtis distance in R (http://www.R-project.org). Nonmetric multidimensional scaling ordination (NMDS) was used to show similarity in genome composition of selected symbiotic bacteria. Differences in overall genome composition, based on COG clusters, among different host feeding strategies, and between the character of symbiotic relationships (primary versus secondary) were assessed using permutational multivariate test implemented in the Adonis function in R (http://www.R-project.org).

Multiple alignments and phylogenetic analyses. Matrices for individual bacterial genera and their outgroups were compiled from the data obtained in this study and the data retrieved from public databases. Complete data sets were aligned using the Mafft algorithm (34) implemented in SeaView (35) software and concatenated into multigene matrices. These contained sequences for the *ftsK*, *fbaA*, *yaeT*, and *spoT* genes of *Arsenophonus*, the *gltA*, *groEL*, *ribC*, and *rpoB* of *Bartonella*, and the *gatB*, *coxA*, *hcpA*, *ftsZ*, and *fbpA* of *Wolbachia*. Individual concatenates were processed in the GBlocks application (36) in order to remove unreliably aligned positions. For the genus *Arsenophonus*, two matrixes were generated: the multigene concatenate described above and, in order to include "*Candidatus* Riesia pediculicola" into the analyses, an incomplete matrix containing the *fbaA* gene present in the extremely reduced genome of this bacterium. The resulting alignments were analyzed using maximum likelihood (ML) and Bayesian inference (BI). The best evolutionary models fitting the data were selected using the program jModelTest (37). ML-based analyses and 100 nonparametric bootstrap replicates were performed in the PhyML program (38). Bayesian analyses were performed with correspondent evolutionary models implemented in MrBayes, version 3.1.2 (39).

Reconstruction of metabolic pathways: B vitamins. Several independent approaches were employed to reliably reconstruct B-vitamin and cofactor pathways of all major bacteria present in *M. ovinus*-"*Ca.* Arsenophonus," "*Ca.* Sodalis," and *B. melophagi*. First, annotated draft genome assemblies were uploaded into Pathway Tools software using its pathological module (40), and individual pathways were compared with species comparisons and manually examined. Second, B-vitamin and cofactor genes from *Escherichia coli* K-12 MG1655 were searched by blastp (E value of 1e−5) against a custom database of all protein coding genes from all three bacteria. Third, putatively missing genes were checked by blasting homologues from *E. coli*, *S. glossinidius*, *Arsenophonus nasoniae*, and *Bartonella schoenbuchensis* species (blastx; E value, 1e−5) against a custom database of all contigs/scaffolds from bacteriome and "*Ca.* Sodalis" assemblies. B-vitamin and cofactor pathways were reconstructed with EcoCyc and KEGG databases as guides (41, 42).

Accession numbers. The "*Ca.* Arsenophonus" draft genome is provided in GenBank format (available at http://users.prf.jcu.cz/novake01/). The "*Ca.* Sodalis" and *B. melophagi* draft genomes are provided as fasta files with all scaffolds/contigs (http://users.prf.jcu.cz/novake01/). All genomes are currently in the process of gap closure and will be submitted to NCBI when closed into a single circular molecule (or several, in case of plasmids). Raw Illumina reads are available in the ENA database under accession number PRJEB9958.

RESULTS

Origin and distribution of the associated bacteria. Screening of the shotgun metagenomic data revealed five microorganisms in the sheep ked microbiota. Phylogenetic analyses showed that four of them belong to the well-known bacterial genera *Arsenophonus*, *Sodalis*, *Bartonella*, and *Wolbachia* (Fig. 1) and that one is the unicellular eukaryote *Trypanosoma melophagium* (see Fig. S1 in the supplemental material). The distribution of the bacteria in the host digestive and reproductive tracts, including overall arrangement of the bacteriome and likely transmission routes the symbionts undertake when passed to the progeny, is presented in Fig. 2 and 3.

The *Arsenophonus* bacterium possesses a number of features also found for some other obligate symbionts. The bacterium was present in all of the examined adults, housed in specialized cells of the host intestine wall (bacteriocytes), forming a compact symbiotic organ (bacteriome) (Fig. 2a to g). Under TEM, the bacteriocytes were tightly packed with pleomorphic populations of the symbiont cells (Fig. 2b). Apart from this typical location, in adult females, these bacteria also occur extracellularly in the lumen of the milk glands (Fig. 3a). This suggests that, similar to *Wigglesworthia* in tsetse flies, this *Arsenophonus* member also uses milk glands for its vertical transmission. Correspondingly, it was never observed in oocytes and early embryos, while its presence was confirmed in the late stages of embryogenesis (see Fig. S2 in the supplemental material). Based on the biological, phylogenetic, and evolutionary traits described in this study, we propose the designation "*Candidatus* Arsenophonus melophagi" (me-lo-phag-i; melophagi, of the genus *Melophagus*).

In contrast to the universally present "*Ca.* Arsenophonus," "*Candidatus* Sodalis melophagi" does not occur in all adult hosts (diagnostic PCR performed on 82 individuals from a single population indicates approximately a 50% prevalence, with 59% of females and 42% of males positive). In the gut, the bacterium resides predominantly outside the bacteriome. However, in contrast to *S. glossinidius*, which has never been described from the bacteriome, "*Ca.* Sodalis" may enter the symbiotic organ. Here, it occupies intercellular space surrounding individual bacteriocytes (Fig. 2e). Similar to "*Ca.* Arsenophonus," "*Ca.* Sodalis" was also detected in the milk glands. However, in addition to the extracellular location, it was also found intracellularly inside the secretory cells (Fig. 3a). Rarely, both bacterial symbionts co-occur in the

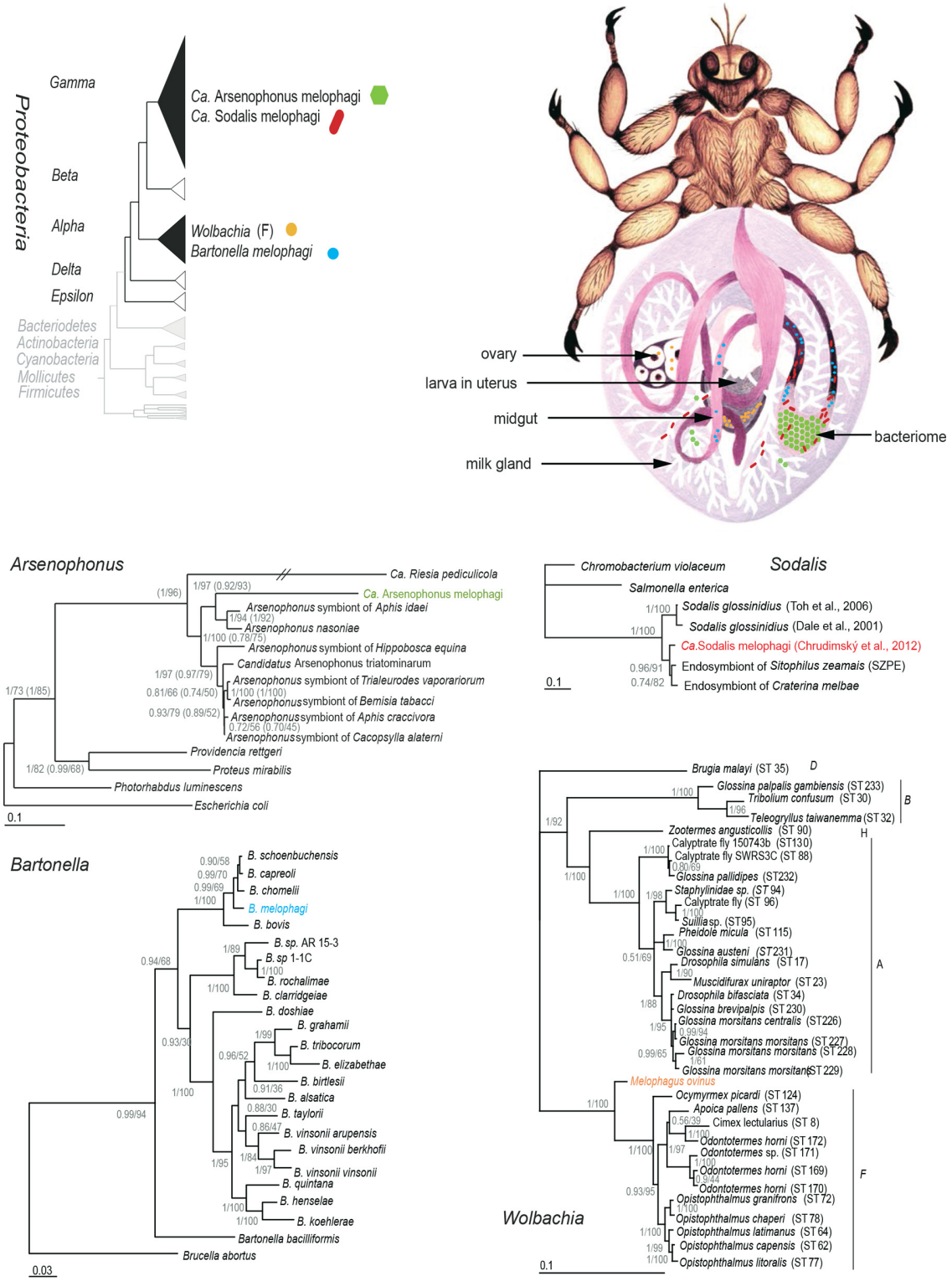


FIG 1 Phylogenetic origins and the overall arrangement of the *Melophagus* symbionts. Phylogenetic position and distribution of the bacteria associated with *Melophagus ovinus* are highlighted as follows: green, "*Ca*. Arsenophonus"; red, "*Ca*. Sodalis"; blue, *B. melophagi*; orange, *Wolbachia*. For *Wolbachia*, phylogeny branches are labeled with host names. Vertical bars and capital letters stand for *Wolbachia* supergroups. Numbers at branch nodes indicate posterior probability and bootstrap values. In *Arsenophonus* phylogeny, numbers in parentheses represent posterior probability and bootstrap values produced by the incomplete matrix with *fbaA* from "*Ca*. Riesia pediculicola" included. The references included are references 19 (Chrudimský et al., 2012), 67 (Dale et al., 2001), and 68 (Toh et al., 2006).

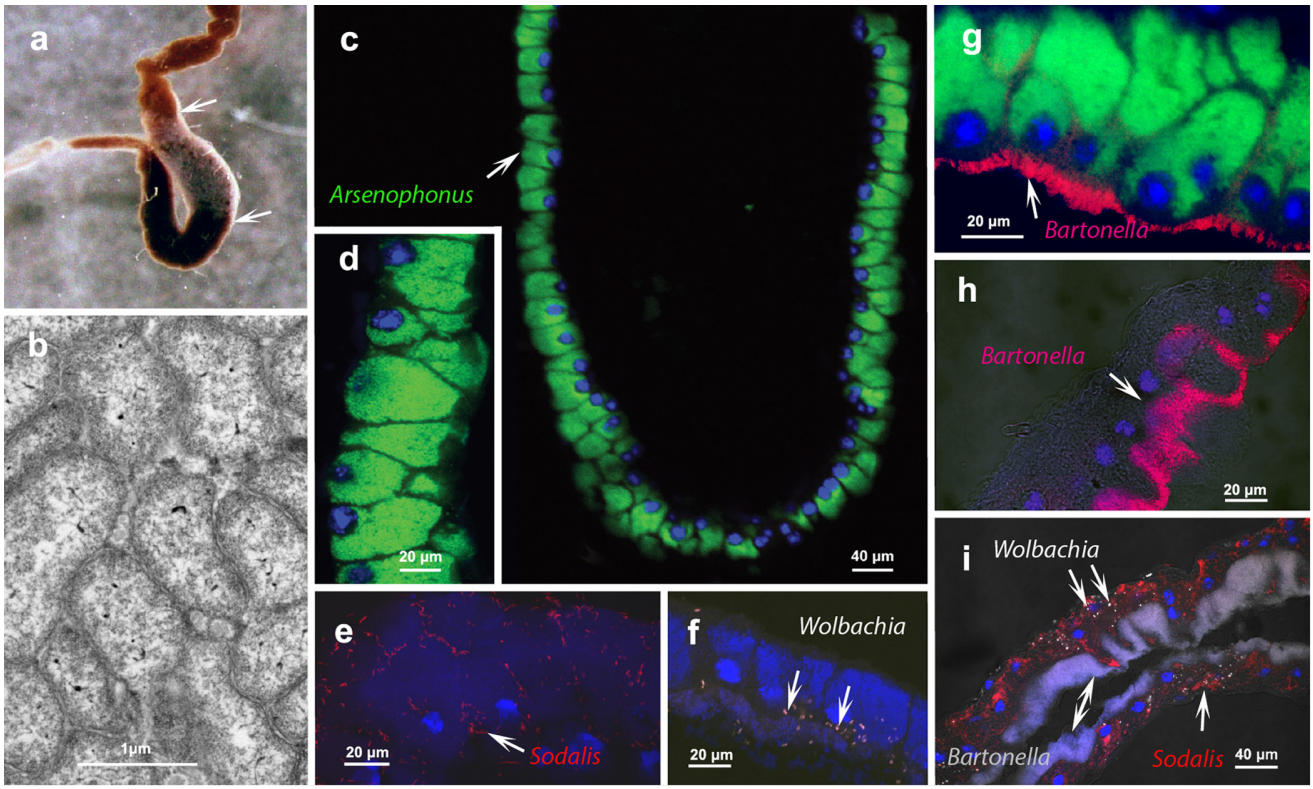


FIG 2 Light microscopy, fluorescence microscopy, and transmission electron microscopy (TEM) of *M. ovinus* gut sections. (a) Dissected midgut section with the symbiotic organ bacteriome designated with white arrows. (b) TEM of the bacteriome packed with "*Ca.* Arsenophonus" bacteria. Visualization Endosymbionts were visualized by 16S rRNA-targeted fluorescent *in situ* hybridization using multiple labeled probes: Flc, Cy3, Cy5, and Texas Red. Arrows and genus names indicate signals for different bacteria. (c, d, and g) "*Ca.* Arsenophonus" (Flc label) bacteria housed in the bacteriome. (e and f) Distribution of "*Ca.* Sodalis" (Cy3) and *Wolbachia* sp. (Cy5) throughout the bacteriome. *B. melophagi* (Texas Red) is present along the bacteriome (g) and in the rest of the midgut (h). (i) "*Ca.* Sodalis" (Cy3) and *Wolbachia* sp. (Cy5) in the midgut section. A nonspecific signal of *B. melophagi* (no probe) cells in the gut lumen is highlighted with a bidirectional arrow in panel i.

gland lumen (see Fig. S2 in the supplemental material). No "*Ca.* Sodalis" cells were detected in oocytes or early embryos.

Wolbachia found in *M. ovinus* branches as a sister lineage to the F supergroup (Fig. 1) and is highly prevalent in the screened population (100% of females and 94% of males positive) (see Materials and Methods for more detail on diagnostic PCR). *Wolbachia* was observed intracellularly in various tissues, including fat bodies (adipocytes) (Fig. 3b), secretory cells of the milk glands (Fig. 3a), and the gut tissue (Fig. 2f and i). Within the bacteriome, *Wolbachia* was scattered in small numbers among bacteriocytes (Fig. 2f). Although present in the secretory cells, *Wolbachia* was never observed in the milk gland lumen. In contrast to "*Ca.* Arsenophonus" and "*Ca.* Sodalis" symbionts, it infected embryos prior to intrauterine development (Fig. 3d). The findings described above distinguish the *Melophagus*-associated *Wolbachia* from the strains found in *Glossina*.

Bartonella melophagi was found in all of the screened individuals and is the only one of the detected bacteria located mostly extracellularly. It was found along the entire gut lumen (Fig. 2g and h) closely attached to the microvilli (see Fig. S1 in the supplemental material). Its abundance is particularly high along the gut wall outside the bacteriome section (Fig. 2h), whereas the cell numbers decrease within the bacteriome (Fig. 2g). Except for the gut, all the other tissues were found to be *Bartonella* free.

Genome comparisons. The shotgun metagenomic data combined with genomic data from "*Ca.* Sodalis" culture (see Materials and Methods) allowed for draft assembly of three bacterial genomes, i.e., "*Ca.* Arsenophonus," *Bartonella melophagi*, and "*Ca.* Sodalis," each with coverage above 100-fold (Table 1). The only bacterium with the coverage too low for further analysis was the symbiont of the genus *Wolbachia*. The sequencing and the assembly strategy we used resulted in genome assemblies truncated in the number of contigs/scaffolds by multiple rRNA operons ("*Ca.* Arsenophonus") and a number of duplicated genes and mobile genetic elements in the genomes of *Bartonella melophagi* and "*Ca.* Sodalis" (Table 1). The high coverage and the genome properties, particularly the number of tRNA genes and retained single-copy (SICO) genes in gammaproteobacterial genomes (see Materials and Methods), imply that the drafts represent the complete gene set. The *Bartonella* genome assembly fully reflects the gene content of the U.S. isolate of *B. melophagi* deposited in the NCBI Assembly database (GCA_000278255), sharing complete gene synteny and ~99% sequence identity in single-copy genome regions. Altogether, the basic genome characteristics listed in Table 2 underpin our microscopic findings, resembling the tsetse fly symbiosis. While the low-GC content and genome size of "*Ca.* Arsenophonus" are in line with its obligate relationship to the host, the con-

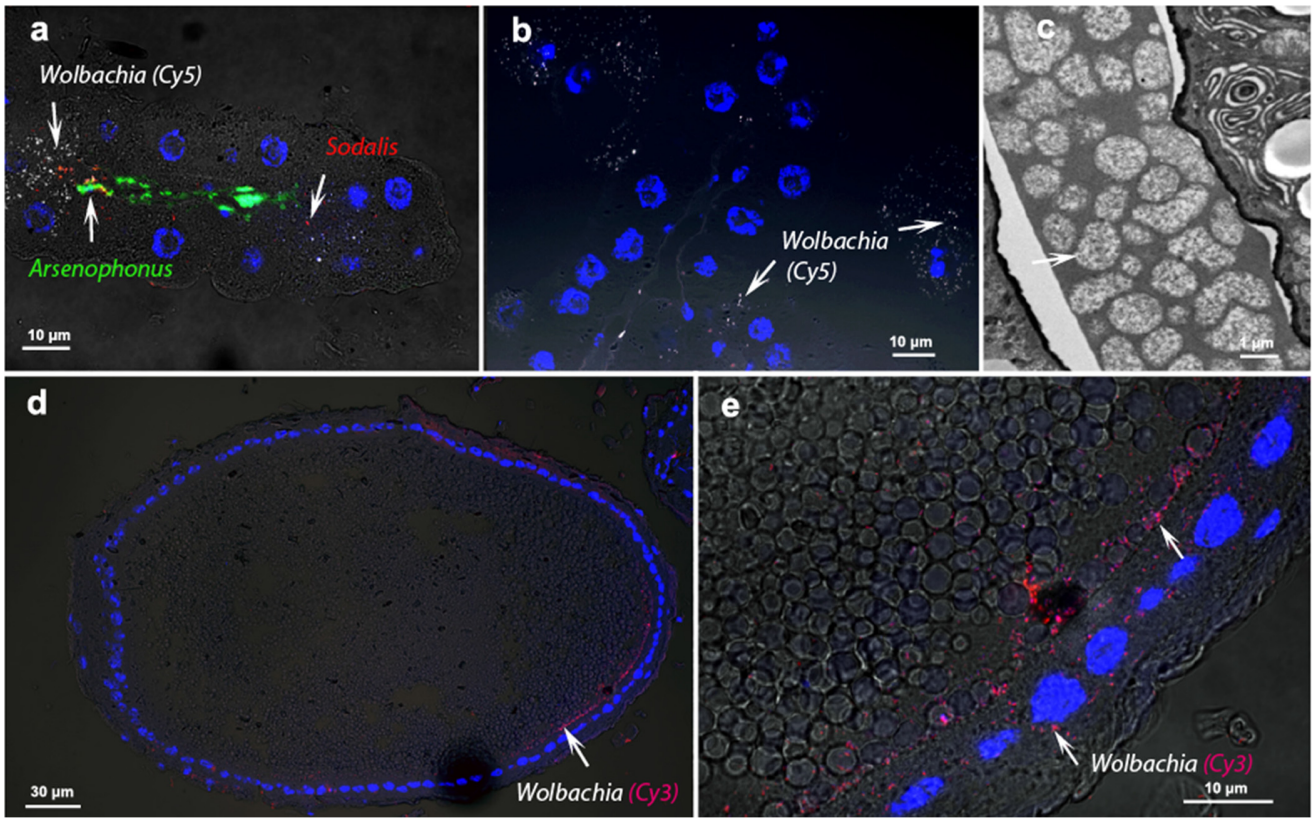


FIG 3 Fluorescence microscopy and transmission electron microscopy (TEM) of *M. ovinus* reproductive tract. Arrows and genus names indicate signals for different bacteria. (a) "*Ca.* Arsenophonus" (Flc label) bacteria present in the milk gland lumen, "*Ca.* Sodalis" (Cy3) in the lumen and intracellularly in the secretory cells of the gland, and *Wolbachia* sp. (Cy5) in milk gland secretory cells. (b) *Wolbachia* sp. in milk gland secretory cells and adipocytes. (c) TEM of the milk gland lumen containing "*Ca.* Arsenophonus" bacteria. (d) *Wolbachia* sp. in developing embryo. (e) Enlarged view of panel d.

siderably larger genome of "*Ca.* Sodalis" and its unbiased base composition suggest a less intimate association.

Compared to other *Arsenophonus* species and the bacterial symbionts from other taxa (see Table S1 in the supplemental material), "*Ca.* Arsenophonus" displays a remarkable degree of similarity with the tsetse fly symbiont *W. glossinidia* in terms of its genome content (Table 2; see also Table S2 in the supplemental material). Both genomes contain similar gene numbers and show similar patterns of present/absent orthologous genes. The ordination plot of overall similarities among 44 symbiotic bacteria, based on the NMDS analysis (Fig. 4), shows remarkably close distance between "*Ca.* Arsenophonus" and both *Wigglesworthia* genomes. As expected, the statistical evaluation of the clustering pattern also showed significant differences ($P = 0.001$) between genome composition of primary and secondary symbionts and among symbionts of hosts with different feeding strategies (i.e., blood-feeding, sap-sucking, and omnivorous insects) (Fig. 4).

Using the assignment of the orthologous clusters to individual

TABLE 1 Genome assembly metrics

Parameter	"*Ca.* Arsenophonus"	*Bartonella melophagi*	"*Ca.* Sodalis"
Approximate contig coverage (no.)	1,360	151	138
Total no. of contigs (no. of scaffolds)	40	74 (67)	616 (236)
Total no. of bases (including unresolved nucleotides)[a]	1,155,312	1,480,425	4,569,254
Total no. of unresolved nucleotides	0	180	423,006
Number of contigs with >1,000 bp	36	40	251
Largest contig (scaffold [bp])	116,563	502,641	153,948 (387.476)
Scaffold N50 (bp)[b]	87,029	381,476	171,607
Scaffold N75 (bp)[b]	33,043	144,087	95,759
GC content (%)	32.2	36.9	50.8
No. of present SICO genes[c]	203	N/A	205

[a] Introduced by the assembly process.
[b] N50 and N75, 50% and 75% of the length of the genome, respectively, is contained in contigs of the specified sizes or larger.
[c] Total number, 205.

TABLE 2 Comparison of basic genome characteristics between symbionts of *Melophagus ovinus* and *Glossina morsitans*

Parameter	*Melophagus ovinus* symbiont		*Glossina morsitans* symbiont	
	"*Ca.* Arsenophonus melophagi"	"*Ca.* Sodalis melophagi"	*Wigglesworthia glossinidia*	*Sodalis glossinidius*
Genome size (Mb)[a]	1.16	4.15	0.72	4.29
GC content (%)	32.2	50.8	25.2	54.5
Coding density (%)	61.3	84.9	87.9	78.4
No. of predicted CDS	765	4,964	678	5,812
tRNA genes (no.)	34	66	35	71
Protein coding genes (no.)	725	4,889	635	5,634
Average CDS length (bp)	926	653	933	579
No. of shared COGs	462	1,453	462	1,453

[a] Excluding unresolved nucleotides introduced by the assembly process.

COG categories (as defined by the KEGG database), the ratios of shared/present/absent orthologs was found to differ considerably across functional categories (see Tables S2 and S3 in the supplemental material). While in some categories (e.g., translation, ribosomal structure, and biogenesis; intracellular trafficking, secretion, and vesicular transport; and nucleotide transport and metabolism) the three genomes share similar subsets of genes, in others they differ considerably. Such a case is best illustrated by the categories replication, recombination, and repair, energy production and conversion, and cell motility. The "*Ca.* Arsenophonus" genome prevails in the first two categories, whereas the genome of *W. glossinidia* possesses a higher number of the genes in the last category. The cell motility category provides an extreme example where *W. glossinidia* possesses 28 genes, but the entire category is missing in "*Ca.* Arsenophonus." A further assignment of the genes within each COG category to particular metabolic pathways shows that the differences between the two symbionts in the gene counts are often due to absence/presence of a metabolic path or its part, rather than to random loss of individual genes (see Table S4 in the supplemental material).

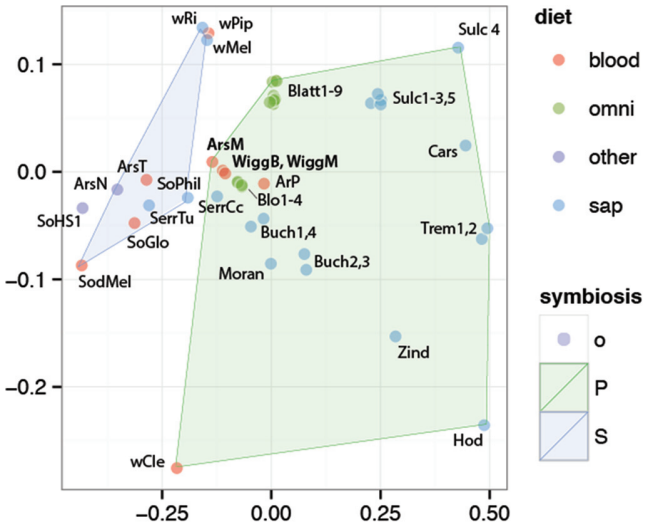


FIG 4 Genome-wide similarity among selected primary and secondary symbionts. The plot (constrained ordination) shows similarities among 44 bacterial genomes based on numbers of orthologous genes transformed into Bray-Curtis distances. ArM, "*Ca.* Arsenophonus melophagi"; WiggB and WiggM, *Wigglesworthia glossinidia* from *Glossina brevipalpis* and *Glossina morsitans*, respectively. A full list of the abbreviations and further information on analyzed genomes is provided in Table S1 in the supplemental material. P and S designate primary and secondary symbionts.

Metabolism: B-vitamin pathways. Because the crucial role of B vitamins produced by symbiotic bacteria in hematophagous systems has been generally accepted and because other possible functions of symbiotic bacteria in blood-feeding arthropods (such as immune system maturation) are impossible to infer based on genome data only, we streamlined the analyses toward B-vitamin biosynthetic pathways. Compared to the *Glossina*-associated symbiotic system, two of the B-vitamin pathways, those for thiamine and pantothenate, were arranged in notably different ways. The genes for the thiamine transporter are present in the obligate symbiont of *M. ovinus*, "*Ca.* Arsenophonus," while the facultative "*Ca.* Sodalis" symbiont codes for its complete biosynthetic pathway. *B. melophagi* is also capable of thiamine biosynthesis but uses an alternative pathway starting from glycine (Fig. 5). In correspondence with the absence of pantothenate biosynthesis genes (*panBCDE*) in the known *Arsenophonus* chromosomes (BioSample accession numbers SAMN02440620, SAMN02603451, and SAMN03078484), "*Ca.* Arsenophonus" also lacks these genes and carries only the genes necessary for synthesis of coenzyme A (CoA) from pantothenate (Fig. 5). On the other hand, the genome reduction in the obligate symbionts of blood-sucking insects led to losses of several identical enzymes in B-vitamin pathways, e.g., *phoA* in folate and *nadABC* in nicotinate pathways (see Fig. S3 in the supplemental material).

DISCUSSION

Biological convergence between the symbioses. The symbiotic system of *Melophagus ovinus* shows a remarkable overall analogy to that described for the related hippoboscoid group, the genus *Glossina*. Compared in more detail, these two systems possess an interesting combination of common features and unique characteristics. While the two obligate mutualists *W. glossinidia* and "*Ca.* Arsenophonus" originated independently from two phylogenetically distant groups, they share several important traits. The most apparent trait is their location in adult hosts within the bacteriomes formed by enlarged cells of the intestinal wall (Fig. 2). Considering their nutritional role, this seems to be a suitable location and has been reported for several other hosts, but it is certainly not the only and inevitable arrangement of symbionts in blood feeders, as documented by a variety of bacteriome structures and locations in sucking lice, bat flies, or bed bugs (43–45). Even more striking is the close resemblance of the transmission modes. Both bacteria, "*Ca.* Arsenophonus" and *W. glossinidia*, constitute two

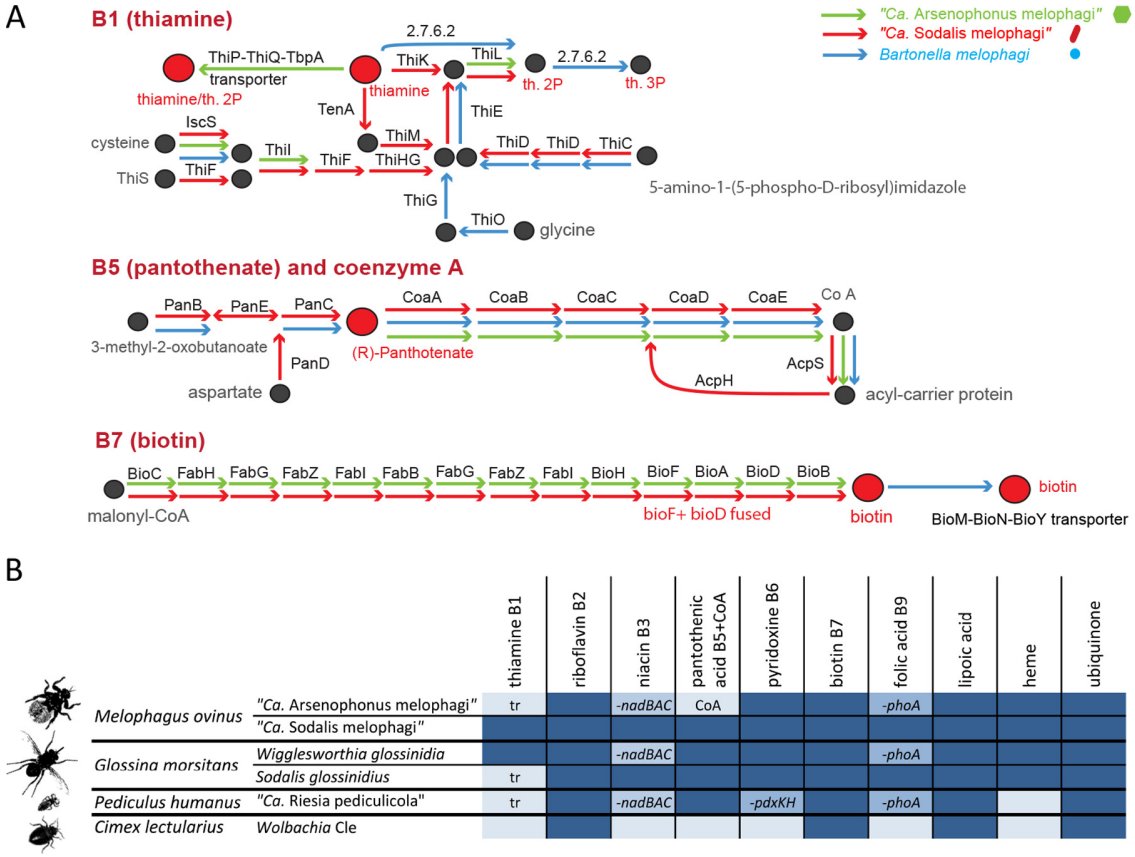


FIG 5 Selected B-vitamin pathways based on draft genomes of *Arsenophonus*, *Sodalis*, and *Bartonella*. (A) The arrows indicate presence of particular genes for different bacteria in *Melophagus ovinus* according to the color legend. (B) The presence of a complete pathway is indicated with the dark blue. Lighter shades are for pathways compromised by missing (−) genes named in the figure. The lightest color indicates corrupted pathways with more than three missing genes. tr, presence of the thiamine ABC transporter.

separate subpopulations in host adult females, one within the bacteriome and the other in the lumen of the milk glands (46–48). A similar arrangement was also reported for the *Arsenophonus*-related symbionts "*Candidatus* Aschnera chinzeii" in the bat fly *Penicillidia jenynsii* (44). It indicates that in these insects the vertical transmission takes place via the accessory organs rather than oocytes, and this mechanism has originated in or been preserved for different groups of bacterial symbionts.

The same transmission path to the developing embryo is also used by the facultative symbionts of the genus *Sodalis*. However, according to our microscope observations, its transmission is not synchronized with that of "*Ca.* Arsenophonus." Rather, "*Ca.* Arsenophonus" seems to colonize the developing embryo prior to "*Ca.* Sodalis." Similarly, the common presence in the host milk gland lumen was only rarely reported for *W. glossinidia* and *S. glossinidius* in Glossinidae (48). This suggests that different mechanisms initiate the vertical transmission of facultative and obligate symbionts in *M. ovinus* or that there is possible competition between them, as proposed previously for the *Glossina* symbionts (48).

The presence of *B. melophagi* along microvilli of the *M. ovinus* midgut indicates possible scavenging of nutrients (e.g., biotin) (from blood digestion and perhaps from the other symbionts (Fig. 5). Although this bacterium, originally cultivated from sheep keds by Noller (49) and described as *Rickettsia melophagi*, was later also cultivated from commercial batches of sheep blood (50), its life cycle is still ambiguous. Halos et al. (51) reported that despite its presence in all *Melophagus* samples collected from a sheep, *B. melophagi* could not be isolated from the sheep itself and explained this by its symbiotic association without transmission to ruminants. Our genomic data do not provide any strong evidence to confirm or rule out this hypothesis. Additional functional data will be needed to infer the exact role of *B. melophagi* in the system.

Glossinidae-Hippoboscidae, a proposed symbiotic model for comparative genomics. While evolving from different ancestors, the two obligate symbionts "*Ca.* Arsenophonus" and *W. glossinidia* have been subject to similar host-mediated influences, reflecting phylogenetic relationships of the two insect groups and their nutritional reliance on vertebrate blood. They thus provide a unique system, potentially useful for distinguishing specific adaptive modifications of the genome from the general degenerative processes. For example, apart from genome reduction, which by itself is hardly surprising as it is a trait generally shared by the obligate symbionts, the two symbionts display high similarity in the pattern of the present/absent orthologous genes (Fig. 4). However, despite this overall similarity, the degree of genomic changes differs across the COG categories (see Tables S2 and S3 in the supplemental material), indicating a possible role of adaptive pro-

cesses. Two patterns may be of particular interest. The first of them is exemplified by the COG category coenzyme transport and metabolism, for which the two hippoboscoid symbionts possess the highest number of genes among all included obligate symbionts. It is interesting that the high gene content of this category is also found in otherwise strongly reduced genomes of "*Ca*. Riesia," a symbiont from sucking lice, and mutualistic *Wolbachia* from bed bugs. This makes the category a promising candidate for the traits possibly important in blood-feeding hosts. A second important pattern is provided by categories where the two symbionts differ considerably from each other in the numbers of present genes. These categories are useful in discriminating the traits that are not associated with the common lifestyle. From this perspective, the difference in the COG category motility is of particular interest. It has been speculated previously that the flagellar system present in the *Wigglesworthia* genome is necessary for transmission of the symbiont via the host milk glands or to enable entry into the larval or pupal bacteriocytes (52). However, in "*Ca*. Arsenophonus," which uses the same transmission route (Fig. 3a and c; see also Fig. S2 in the supplemental material), this COG category is entirely absent. While this finding does not exclude use of the flagella for the transmission in *Wigglesworthia*, it at least indicates that this motility apparatus is not necessary for such a transmission mode. Another striking example for the "*Ca*. Arsenophonus"-*W. glossinidia* comparison is the category energy production and conversion, where the difference of 18 genes is largely due to the presence of an Rnf/Nqr complex and ferredoxin in the "*Ca*. Arsenophonus" genome and its absence in *W. glossinidia*. Similar differences between the two symbionts are also found in some other COG categories, e.g., defense mechanisms, signal transduction mechanisms, and replication, recombination, and repair (Tables S2 and S3 in the supplemental material). The number of genes in the last category is much higher in the "*Ca*. Arsenophonus" genome (as well as the set of genes with unknown functions). This is in good agreement with the perception of "*Ca*. Arsenophonus" as a less degenerate symbiont.

To interpret these observations on a rigorous comparative basis, it will be important to distinguish the variability due to different physiological demands (i.e., adaptive processes) from that reflecting evolutionary age of a given symbiont and/or stochastic processes. This can only been achieved in a broader phylogenetic framework of the hippoboscid-*Arsenophonus* associations using large data sets and proper statistical analysis. The polyphyly of the Hippoboscidae-*Arsenophonus* symbiosis, revealed by 16S rRNA gene- and MLST-based phylogenies (21, 53), suggests its multiple independent origins. This could explain why the genome of "*Ca*. Arsenophonus" is slightly larger than that of *W. glossinidia* if "*Ca*. Arsenophonus" is considered a younger symbiont which originated more recently from a free-living bacterium (possibly replacing other, more ancient symbiont of hippoboscids). In theory the higher number of "*Ca*. Arsenophonus" genes could also be due, at least partially, to gene acquisitions via horizontal gene transmission and/or gene duplications. A simple check showed that a majority of the 725 "*Ca*. Arsenophonus" protein-coding genes returned the best or second-best BLAST hits corresponding to the genus *Arsenophonus* (664) or the closely related genera *Providencia*, *Proteus*, *Photorhabdus*, and *Xenorhabdus* (16). Eighteen hypothetical proteins did not return any hit; for 27 of the remaining genes phylogenetic affinity is less clear and will require more detailed phylogenetic analysis.

Implications for B-vitamin and cofactor biosynthesis in blood-sucking insects. Biosynthesis of B vitamins is traditionally considered a core aspect of the associations between bacteria and blood-sucking insects (13, 43, 54, 55). In our comparison, the arrangement of thiamine biosynthesis and transport presents one of the unexpected metabolic differences between the tsetse fly and louse fly symbiotic systems. In tsetse flies, the genome of *W. glossinidia* codes for the thiamine biosynthetic pathway, and *S. glossinidius* possesses only its ABC transporter (52, 56, 57). In *M. ovinus*, it is completely reversed. Our current data are not sufficient to infer the exact role of thiamine in the system; rather they allow for several not mutually exclusive hypotheses: (i) thiamine is provided to *M. ovinus* and/or "*Ca*. Arsenophonus"by "*Ca*. Sodalis" and/or *B. melophagi*; (ii) "*Ca*. Sodalis" and *B. melophagi* use thiamine solely for their own metabolism, but it is also available from blood *for* "*Ca*. Arsenophonus"; (iii) thiamine supplemented by the symbionts is beneficial but not necessary for host survival (during the whole life cycle or only in certain life stages). Similar hypothetical scenarios may apply to pantothenate biosynthesis: either it is provided by "*Ca*. Sodalis," or it is available from other sources, i.e., sheep blood. This arrangement of the B-vitamin biosynthesis genes and the implied hypotheses indicate that the nutritional interactions between bacteria and blood-feeding insects may be more complex (less determined) than usually supposed. From this perspective, it is interesting that the only other obligate symbiont known to code for thiamine transporter is another member of the *Arsenophonus* clade, "*Ca*. Riesia pediculicola" (58). Since "*Ca*. Riesia" is an exclusive symbiont in human lice, thiamine is likely available from blood and needed by the symbionts for their own metabolism. This uncertainty further highlights our limited knowledge on symbiotic systems of blood feeders. Unlike the symbiotic associations in sap-sucking insects (where symbionts provide their hosts with well-known sets of essential amino acids [59]), it is still quite unclear which B vitamins are provided to blood-sucking hosts and which are needed only for symbionts, with the exception of pyridoxine and thiamine provisioning in the tsetse fly model (11, 14, 15).

Diversity of microorganisms associated with Hippoboscoidea. The system presented here, i.e., the bacteria regularly associated with *M. ovinus*, is an example of the diverse symbiotic associations known from other hippoboscoids. As noted above, the presence of *Arsenophonus* bacteria in different Hippoboscidae members is likely due to multiple origins of the symbiosis or perhaps to recurrent replacements of the obligate symbionts by facultative *Arsenophonus* bacteria (21). The affinity to *Arsenophonus* symbionts is also well demonstrated in a related hippoboscoid group, the bat flies of the family Nycteribiidae (44, 60). Similarly, two other bacterial genera reported here from *M. ovinus*, namely, *Bartonella* and *Wolbachia*, have also been previously identified in other Hippoboscoidea (61–65). *Melophagus ovinus* is, however, the only host known so far with fixed infection of *Bartonella* (51; also this study). Considering this diversity and the abundance of the hippoboscoid-associated symbionts, together with the results presented here for a single hippoboscoid species, this group seems to be a valuable model for further comparative analyses of the insect-bacterium evolution.

ACKNOWLEDGMENTS

This work was supported by the Grant Agency of the Czech Republic (grant 13-01878S to V.H.) and the project Postdoc USB (registration no.

CZ.1.07/2.3.00/30.0006) realized through the EU Education for Competitiveness Operational Programme.

We certify that we have no involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in the manuscript.

REFERENCES

1. **Baumann P.** 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol **59:**155–189. http://dx.doi.org/10.1146/annurev.micro.59.030804.121041.
2. **McCutcheon JP, Moran NA.** 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol **10:**13–26. http://dx.doi.org/10.1038/nrmicro2670.
3. **Moran NA, McCutcheon JP, Nakabachi A.** 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet **42:**165–190. http://dx.doi.org/10.1146/annurev.genet.41.110306.130119.
4. **Petersen FT, Meier R, Kutty SN, Wiegmann BM.** 2007. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. Mol Phylogenet Evol **45:**111–122. http://dx.doi.org/10.1016/j.ympev.2007.04.023.
5. **Meier R, Kotrba M, Ferrar P.** 1999. Ovoviviparity and viviparity in Diptera. Biol Rev **74:**199–258. http://dx.doi.org/10.1017/S0006323199005320.
6. **Lenoble BJ, Denlinger DL.** 1982. The milk gland of the sheep ked, *Melophagus ovinus*: a comparison with *Glossina*. J Insect Physiol **28:**165–172. http://dx.doi.org/10.1016/0022-1910(82)90124-X.
7. **Aksoy S.** 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of mycetocyte associated, primary endosymbionts of tsetse flies. Int J Syst Bacteriol **45:**848–851. http://dx.doi.org/10.1099/00207713-45-4-848.
8. **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. http://dx.doi.org/10.1046/j.1365-2583.1999.810125.x.
9. **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. http://dx.doi.org/10.1099/00207713-49-1-267.
10. **Aksoy S, Chen X, Hypša V.** 1997. Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). Insect Mol Biol **6:**183–190. http://dx.doi.org/10.1111/j.1365-2583.1997.tb00086.x.
11. **Snyder AK, Deberry JW, Runyen-Janecky L, Rio RV.** 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc Biol Sci **277:**2389–2397. http://dx.doi.org/10.1098/rspb.2010.0364.
12. **Nogge G.** 1976. Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. Experientia **32:**995–996. http://dx.doi.org/10.1007/BF01933932.
13. **Nogge G.** 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology **82:**101–104.
14. **Snyder AK, McLain C, Rio RV.** 2012. The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density via exogenous nutrient provisioning. Appl Environ Microbiol **78:**7792–7797. http://dx.doi.org/10.1128/AEM.02052-12.
15. **Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S.** 2014. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. Appl Environ Microbiol **80:**5844–5853. http://dx.doi.org/10.1128/AEM.01150-14.
16. **Doudoumis V, Alam U, Aksoy E, Abd-Alla AMM, Tsiamis G, Brelsfoard C, Aksoy S, Bourtzis K.** 2013. Tsetse-*Wolbachia* symbiosis: comes of age and has great potential for pest and disease control. J Invertebr Pathol **112:**S94–S103. http://dx.doi.org/10.1016/j.jip.2012.05.010.
17. **Lindh J, Lehane M.** 2011. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie Van Leeuwenhoek **99:**711–720. http://dx.doi.org/10.1007/s10482-010-9546-x.
18. **Aksoy E, Telleria EL, Echodu R, Wu Y, Okedi LM, Weiss BL, Aksoy S, Caccone A.** 2014. Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. Appl Environ Microbiol **80:**4301–4312. http://dx.doi.org/10.1128/AEM.00079-14.
19. **Chrudimský T, Husník F, Nováková E, Hypsa V.** 2012. *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. PLoS One **7:**e40354. http://dx.doi.org/10.1371/journal.pone.0040354.
20. **Nováková E, Hypša V.** 2007. A new *Sodalis* lineage from blood-sucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiol Lett **269:**131–135. http://dx.doi.org/10.1111/j.1574-6968.2006.00620.x.
21. **Nováková E, Hypsa V, Moran NA.** 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol **9:**143. http://dx.doi.org/10.1186/1471-2180-9-143.
22. **Small RW.** 2005. A review of *Melophagus ovinus* (L.), the sheep ked. Vet Parasitol **130:**141–155. http://dx.doi.org/10.1016/j.vetpar.2005.03.005.
23. **Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA.** 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol **56:**1919–1925.
24. **Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH.** 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol **72:**7098–7110. http://dx.doi.org/10.1128/AEM.00731-06.
25. **O'Neill SL, Giordano R, Colbert AME, Karr TL, Robertson HM.** 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci U S A **89:**2699–2702. http://dx.doi.org/10.1073/pnas.89.7.2699.
26. **Norman AF, Regnery R, Jameson P, Greene C, Krause DC.** 1995. Differentiation of *Bartonella*-like isolates at the species level by PCR-restriction fragment length polymorphism in the citrate synthase gene. J Clin Microbiol **33:**1797–1803.
27. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
28. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
29. **Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC.** 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res **40:**D115–D122. http://dx.doi.org/10.1093/nar/gkr1044.
30. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
31. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
32. **Lerat E, Daubin V, Moran NA.** 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the γ-*Proteobacteria*. PLoS Biol **1:**e19.
33. **Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH.** 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res **30:**281–283. http://dx.doi.org/10.1093/nar/30.1.281.
34. **Katoh K, Standley DM.** 2014. MAFFT: iterative refinement and additional methods. Methods Mol Biol **1079:**131–146. http://dx.doi.org/10.1007/978-1-62703-646-7_8.
35. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221–224. http://dx.doi.org/10.1093/molbev/msp259.
36. **Castresana J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17:**540–552. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334.
37. **Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods **9:**772–772. http://dx.doi.org/10.1038/nmeth.2109.
38. **Galtier N, Gouy M.** 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol **15:**871–879. http://dx.doi.org/10.1093/oxfordjournals.molbev.a025991.
39. **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic

inference under mixed models. Bioinformatics **19:**1572–1574. http://dx.doi.org/10.1093/bioinformatics/btg180.
40. **Dale JM, Popescu L, Karp PD.** 2010. Machine learning methods for metabolic pathway prediction. BMC Bioinformatics **11:**15. http://dx.doi.org/10.1186/1471-2105-11-15.
41. **Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M.** 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res **40:**D109–D114. http://dx.doi.org/10.1093/nar/gkr988.
42. **Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroeder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD.** 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res **41:**D605–D612. http://dx.doi.org/10.1093/nar/gks1027.
43. **Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T.** 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A **107:**769–774. http://dx.doi.org/10.1073/pnas.0911476107.
44. **Hosokawa T, Nikoh N, Koga R, Sato M, Tanahashi M, Meng XY, Fukatsu T.** 2012. Reductive genome evolution, host-symbiont cospeciation and uterine transmission of endosymbiotic bacteria in bat flies. ISME J **6:**577–587. http://dx.doi.org/10.1038/ismej.2011.125.
45. **Ries E.** 1931. Die Symbiose der Läuse und Federlinge. Zeitschrift für Morphologie und Ökologie der Tiere **20:**233–367.
46. **Aksoy S, Pourhosseeini AA, Chow A.** 1995. Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to *Enterobacteriaceae*. Insect Mol Biol **4:**5–22.
47. **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. http://dx.doi.org/10.1016/j.jinsphys.2008.06.008.
48. **Balmand S, Lohs C, Aksoy S, Heddi A.** 2013. Tissue distribution and transmission routes for the tsetse fly endosymbionts. J Invertebr Pathol **112:**S116–S122. http://dx.doi.org/10.1016/j.jip.2012.04.002.
49. **Noller W.** 1916. Beitrag zur Flecktyphusubertragung durch Lause. Klin Wochenschr **28:**778.
50. **Bemis DA, Kania SA.** 2007. Isolation of *Bartonella* sp from sheep blood. Emerg Infect Dis **13:**1565–1567. http://dx.doi.org/10.3201/eid1310.070570.
51. **Halos L, Jamal T, Maillard R, Girard B, Guillot J, Chomel B, Vayssier-Taussat M, Boulouis HJ.** 2004. Role of Hippoboscidae flies as potential vectors of *Bartonella* spp. infecting wild and domestic ruminants. Appl Environ Microbiol **70:**6302–6305. http://dx.doi.org/10.1128/AEM.70.10.6302-6305.2004.
52. **Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S.** 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet **32:**402–407. http://dx.doi.org/10.1038/ng986.
53. **Duron O, Schneppat UE, Berthomieu A, Goodman SM, Droz B, Paupy C, Obame Nkoghe J, Rahola N, Tortosa P.** 2014. Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. Mol Ecol **23:**2105–2117. http://dx.doi.org/10.1111/mec.12704.
54. **Puchta O.** 1955. Experimentelle Untersuchungen uber die Bedeutung der symbiose der Kleiderlaus Pediculus vestimenti Burm. Z Parastenkd **17:**1–40. (In German.)
55. **Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T.** 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc Natl Acad Sci U S A **111:**10257–10262. http://dx.doi.org/10.1073/pnas.1409284111.
56. **Belda E, Moya A, Bentley S, Silva FJ.** 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics **11:**449. http://dx.doi.org/10.1186/1471-2164-11-449.
57. **Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, Aksoy S.** 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res **16:**149–156.
58. **Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li H-M, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, et al.** 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A **107:**12168–12173. http://dx.doi.org/10.1073/pnas.1003379107.
59. **Hansen AK, Moran NA.** 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. Mol Ecol **23:**1473–1496. http://dx.doi.org/10.1111/mec.12421.
60. **Morse SF, Bush SE, Patterson BD, Dick CW, Gruwell ME, Dittmar K.** 2013. Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). Appl Environ Microbiol **79:**2952–2961. http://dx.doi.org/10.1128/AEM.03814-12.
61. **Morse SF, Olival KJ, Kosoy M, Billeter S, Patterson BD, Dick CW, Dittmar K.** 2012. Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). Infect Genet Evol **12:**1717–1723. http://dx.doi.org/10.1016/j.meegid.2012.06.009.
62. **Billeter SA, Hayman DTS, Peel AJ, Baker K, Wood JLN, Cunningham A, Suu-Ire R, Dittmar K, Kosoy MY.** 2012. *Bartonella* species in bat flies (Diptera: Nycteribiidae) from western Africa. Parasitology **139:**324–329. http://dx.doi.org/10.1017/S0031182011002113.
63. **Reeves WK, Loftis AD, Gore JA, Dasch GA.** 2005. Molecular evidence for novel *Bartonella* species in *Trichobius major* (Diptera: Streblidae) and *Cimex adjunctus* (Hemiptera: Cimicidae) from two southeastern bat caves, USA. J Vector Ecol **30:**339–341.
64. **Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, O'Neill SL, Aksoy S.** 2000. Tissue distribution and prevalence of *Wolbachia infections in tsetse flies*, *Glossina* spp. Med Vet Entomol **14:**44–50. http://dx.doi.org/10.1046/j.1365-2915.2000.00202.x.
65. **Doudoumis V, Tsiamis G, Wamwiri F, Brelsfoard C, Alam U, Aksoy E, Dalaperas S, Abd-Alla A, Ouma J, Takac P, Aksoy S, Bourtzis K.** 2012. Detection and characterization of *Wolbachia* infections in laboratory and natural populations of different species of tsetse flies (genus *Glossina*). BMC Microbiol **12:**S3. http://dx.doi.org/10.1186/1471-2180-12-S1-S3.
66. Reference deleted.
67. **Dale C, Young SA, Haydon DT, Welburn SC.** 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci U S A **98:**1883–1888. http://dx.doi.org/10.1073/pnas.98.4.1883.
68. **Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S.** 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res **16:**149–156. http://dx.doi.org/10.1101/gr.4106106.

Paper 6:

Nováková, E., & Hypša, V. (2007). A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS microbiology letters*, *269*(1), 131-135.

A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*

Eva Nováková & Václav Hypša

Faculty of Biological Sciences and Institute of Parasitology, České Budějovice, Czech Republic

Correspondence: Václav Hypša, Institute of Parasitology AS CR, Branišovská 31, 370 05 České Budějovice, Czech Republic. Tel.: +420 38 7775441; fax: +420 38 5310388; e-mail: vacatko@paru.cas.cz

Received 8 November 2006; revised 13 December 2006; accepted 13 December 2006. First published online 15 January 2007.

DOI:10.1111/j.1574-6968.2006.00620.x

Editor: Marco Soria

Keywords
Sodalis; symbiont; molecular phylogeny; 16S rRNA gene; type three secretion system; *Craterina*.

Abstract

Symbiotic bacterium closely related to the secondary symbiont of tsetse flies, *Sodalis glossinidius*, has been described from the bloodsucking fly *Craterina melbae*. Phylogenetic analysis of two genes, 16S rRNA gene and component of type three secretion system, placed the bacterium closer to the *Sitophilus*-derived branch of *Sodalis* than to the tsetse symbionts. This indicates that the *Craterina*-derived lineage of *Sodalis* originated independent of the tsetse flies symbionts and documents the capability of *Sodalis* bacteria either to switch between different host groups or to establish the symbiosis by several independent events.

Introduction

The symbiotic bacterium *Sodalis glossinidius* (Dale & Maudlin, 1999) was originally reported as one of three phylogenetically distant bacteria inhabiting tissues of tsetse flies (Aksoy *et al*., 1997). The other two bacteria are the mutualistic *Wigglesworthia glossinidia* (Aksoy, 1995) and the broadly distributed parasite of the genus *Wolbachia* (O'Neill *et al*., 1993). Since its discovery, *Sodalis glossinidius* has become an important subject of a symbiosis-centered research due to several favorable circumstances. Primarily, it attracted attention as a companion of an important bloodsucking insect and was recognized as a possible factor influencing the vector competence/capacity (Maudlin & Ellis, 1985). While this view was later opposed by several authors (Moloo & Shaw, 1989; Geiger *et al*., 2005b), the research of *Sodalis glossinidius* further accelerated after Dale *et al*. (2001) demonstrated that it utilizes the machinery of the type three secretion system (TTSS) to enter the host cell. Owing to this finding and the fact that *Sodalis* could be maintained in *in vitro* laboratory cultures (Welburn *et al*., 1987), *Sodalis glossinidius* became a model organism, particularly in studying host cell invasion in an early stage of the symbiosis evolution. As a result of this attention, the complete genome of *Sodalis glossinidius* has recently been sequenced and analyzed (Toh *et al*., 2006). The preliminary analysis revealed a number of genomic changes associated with the adoption of a symbiotic lifestyle and further stressed the significance of *Sodalis* as a model of evolutionary transition from free-living to symbiotic bacteria. However, genetic characterization of a symbiont and the elucidation of its molecular machineries only represent one facet of the evolutionary picture. To understand fully the biology and the evolutionary potential of such a bacterium, it is important to know its distribution among various hosts, the routes of interspecific transmission and the modes of symbiosis it can adopt in different hosts. This can be well demonstrated by the finding that *Sodalis*-related bacteria, associated with the weevils of the genus *Sitophilus*, display features typical for mutualistic primary symbionts (Heddi *et al*., 1998; Nardon *et al*., 2002; Lefevre *et al*., 2004). Such a phylogenetic-distribution pattern indicates that *Sodalis* may be capable of both the horizontal transfers between different hosts, and the long-term coevolution associated with establishment of a mutualistic relation. This makes the bacterium a potentially excellent model for the study of the changes and trends accompanying the adaptation to different symbiotic modes. Unfortunately, despite the considerable number of symbiotic bacteria described annually from various insects, no other

closely related bacterium has been identified so far. In this study, a new lineage is reported from a bloodsucking relative of tsetse flies: the hippoboscoid species *Craterina melbae* (Diptera; *Hippoboscidae*). Based on two different genes, it is shown that this symbiont has been acquired independent of the *S. glossinidius* inhabiting the tsetse flies.

Materials and methods

The sample of *Craterina melbae* was provided by Pierre Bize, University of Lausanne, Switzerland. Total DNA was isolated from tissue of three adult insects using the DNEasy Tissue Kit (QIAGEN) and used with three different sets of primers for the PCR amplification. Primer pair F40: 5′-GCGGCAAGCCTAACACAT-3′ and R1060: 5′-CTTAACC CAACATTTCTCAACACGAG-3′ was designed to amplify a 1200 bp of 16S rRNA gene; the primer SodF 5′-ACCGCA TAACGTCGCAAGACC-3′ was designed to amplify together with R1060 *c.* 1000 bp long fragment of 16S rRNA gene from *Sodalis* and related bacteria, and the primers SpaPQRf: 5′-ATGATGATGATGAGCCCG-3′ and SpaPQRr: 5′-AGCC CATGCATAACCCAAAA-3′ were adopted from Dale *et al.* (2001) and used to amplify components of the TTSS. The identity and preliminary phylogenetic position of the obtained sequences were checked by BLAST Search, NCBI (http://www.ncbi.nlm.nih.gov). The phylogenetic relationships of both genes, the 16S rRNA gene and the Spa component, were further analyzed in a broader taxonomic context. For 16S rRNA gene, an additional 33 sequences were retrieved from the GenBank (NCBI); they included all available *Sodalis* 16S rRNA gene sequences and 13 other symbiotic and free-living bacteria as outgroups (Fig 1a). Owing to a high similarity of the sequences, the alignment was unequivocal and did not require any testing for aligning parameters. The matrix was designed manually in BIOEDIT program (Hall, 1999) and the few ambiguous positions were removed before the analysis. Phylogenetic analysis was performed in PAUP* (Swofford, 1998) by maximum parsimony (MP) and maximum likelihood (ML) methods using the TBR swapping procedure with 30 replicates of random sequence addition. The MP analysis was performed under the transition: transversion ratio set to 1 : 1, 2 : 1 and 3 : 1, and the strict consensus of all trees was computed. The bootstrap support was obtained by 1000 replicates of MP analysis with Ts/Tv set to 1 : 2. The evolutionary model for ML analysis, was determined in Modeltest 3.6 (Posada & Crandall, 1998). To analyze the Spa genes, the sequences were aligned with those from *Sodalis glossinidius*, *Sitophilus zeamais*-symbiont and two outgroups represented by *Chromobacterium violaceum* and *Salmonella enterica*. Two regions of the Spa fragment displayed homology sufficient for an alignment and meaningful phylogenetic analysis. They included partial SpaP (84 bp) and SpaR (231 bp) genes extracted from the following GenBank sequences: AE016825 (*Chromobacterium violaceum*), AF306650 (*Sodalis glossinidius*), AF426456 (symbiont of *Sitophilus zeamais*), and AL627276 (*Salmonella enterica*). The matrix was analyzed by an exhaustive search for the MP tree and examined by eye for the potential molecular synapomorphies. Within the *Sodalis* group, the genetic distances for all genes were calculated in PAUP* using the Jukes–Cantor model.

The sequences amplified from *Craterina melbae* were deposited in GenBank under the accession numbers EF174495 (16S rRNA gene) and EF174496 (Spa genes).

Results and discussion

Phylogenetic analysis of the 16S rRNA gene, performed together with the two previously described *Sodalis* lineages (inhabiting tsetse flies and weevils), revealed that the *Craterina*-derived bacterium is more closely related to the *Sitophilus* symbionts than to *Sodalis glossinidius* (Fig. 1a). An identical arrangement, although within a taxonomically more restricted sample, was obtained when the SpaP and SpaR genes were analyzed, either as two independent data sets or within a single combined matrix (Fig. 1b). In all of the analyses, the position of *Craterina*-derived bacterium as a sister taxon of weevil symbionts has been unequivocal and no other arrangement was retrieved. Moreover, this position was also well supported by the tree parameters, the distances within the *Sodalis* group (Table 1) and several clear molecular synapomorphies within the Spa genes (including a unique and relatively long motif FSGSIV within the SpaR region; Fig. 2). This phylogenetic picture indicates that the *Sodalis* bacterium in *Craterina* has not been inherited from a common *Glossinidae–Hippoboscidae* ancestor, but was acquired by an independent event. Such a result is far from unexpected as the low 16S rRNA gene diversity within *Sodalis glossinidius* has been considered an evidence of multiple symbiont acquisition even within the tsetse flies (Aksoy *et al.*, 1997). On the other hand, the monophyly of *Sodalis glossinidius* isolates is difficult to reconcile with such a multiple-acquisition scenario. Moreover, a recently performed amplified fragment length polymorphism (AFLP) analysis of *Sodalis glossinidius* from two different species of tsetse flies showed that they constitute two genetically distinct host-specific clades (Geiger *et al.*, 2005a). Owing to this contrast between the *Sodalis glossinidius* monophyly and genetic structure on the one hand and low overall DNA diversity on the other, the process underlying current distribution of *Sodalis glossinidius* has never been satisfactorily elucidated. As a possible solution for this puzzle, Aksoy *et al.* (1997) suggested a physiological constraint preventing *Sodalis gossinidius* from invading other insect groups once it adopted the transmission route via the 'milk glands'. The finding, presented in this paper, of a clearly independent

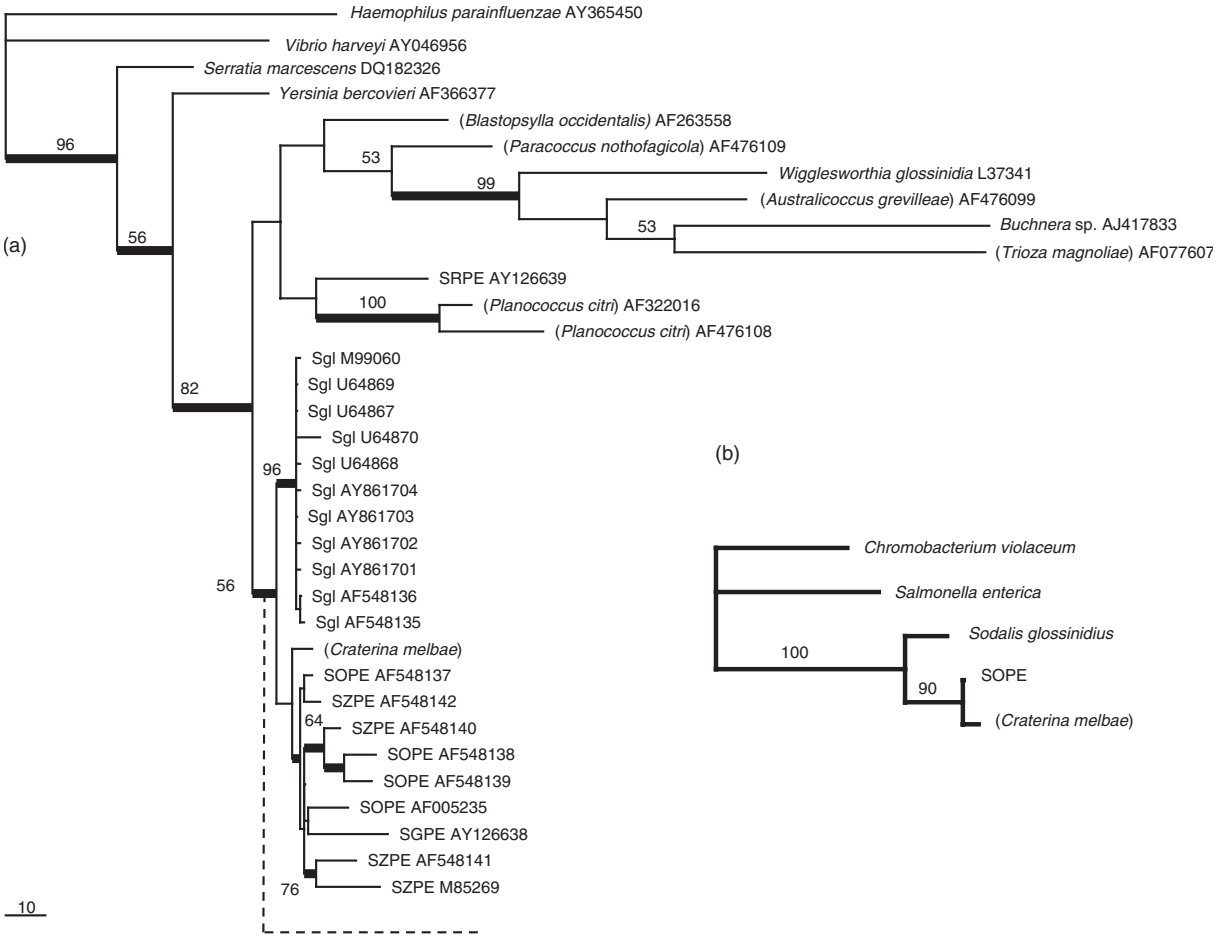


Fig. 1. (a) One of the 38 MP trees obtained by the analysis of 16S rRNA gene matrix under different Ts/Tv ratios (CI of the trees varied from 0.6 to 0.63). The consensus of all trees and compatible with ML analysis is designated by bold lines. The names of host taxa (in brackets) or the bacteria are provided together with their sequence accession numbers. The dashed line indicates an alternative position of the *Sitophilus rugicollis* symbiont as obtained under the nonhomogeneous model. The *Sitophilus*-derived symbionts are labeled *SOPE*, *SGPE*, *SZPE* and *SRPE* for *Sitophilus oryzae*, *Sitophilus granarius*, *Sitophilus zeamais* and *Sitophilus rugicollis*, respecively; Sgl, *Sodalis glossinidius*; (b) the MP tree of SpaP+SpaQ genes retrieved by an exhaustive search (CI = 0.99). In both trees, bootstrap values higher 50% are printed at the nodes.

lineage of *Sodalis* in another hippoboscoid insect, demonstrates the general capability of this bacterium either to switch between different host groups or to establish the symbiosis by several independent events.

Previously, Lefévre *et al.* (2004) hypothesized that the *Sodalis glossinidius* had been acquired by horizontal transfer from *Dryophthoridae* weevils to tsetse flies. They based this suggestion on the intermediate position of *Sodalis glossinidius* between the symbiont of *Sitophilus rugicollis* and the cluster encompassing the symbionts from all other *Sitophilus* spp. When the relationships among the '*Sodalis*' lineages were investigated in a different context of symbiotic and free-living taxa by standard MP and ML methods, a different topology was obtained, with the *Sitophilus rugicollis* symbiont forming a distant lineage independent of other *Sitophilus* symbionts (Fig. 1a). This arrangement can be easily integrated into the general view of several symbiont replacements within the *Dryophthoridae* (Lefevre *et al.*, 2004). Clearly, the problem of the *Sitophilus rugicollis* symbiont rests in the characteristic of its 16S rRNA gene. Compared with the typical *Sodalis* 16S rRNA gene, the sequence of the *Sitophilus rugicollis* symbiont is highly aberrant. It rather resembles those known from most of the P-symbionts, including the *Dryophthoridae* R-clade reported by Lefevre *et al.* (2004), i.e. the sequences typical with a considerable change of AT/GC ratio. As this phenomenon can strongly influence the phylogenetic picture, the application of algorithms dealing with such homogeneities has been suggested in several studies (Galtier & Gouy, 1995, 1998; Lefevre *et al.*, 2004; Herbeck *et al.*, 2005). Indeed, when the

Table 1. Distances calculated by Jukes–Cantor model for the sequences within and between the clades of *Sodalis* group (see Fig. 1 for list of sequences)

	Glossina	*Sitophilus*	*Glossina–Sitophilus*	*Craterina–Glossina*	*Craterina–Sitophilus*
16S	0–0.8	0.7–3.6	1.3–3.5	1.4–2.0	1.0–3.0
SpaP	–	–	16.7	16.7	5.1
SpaQ	–	–	16.3	15.9	0.3
SpaR	–	–	26.7	27.4	3.5

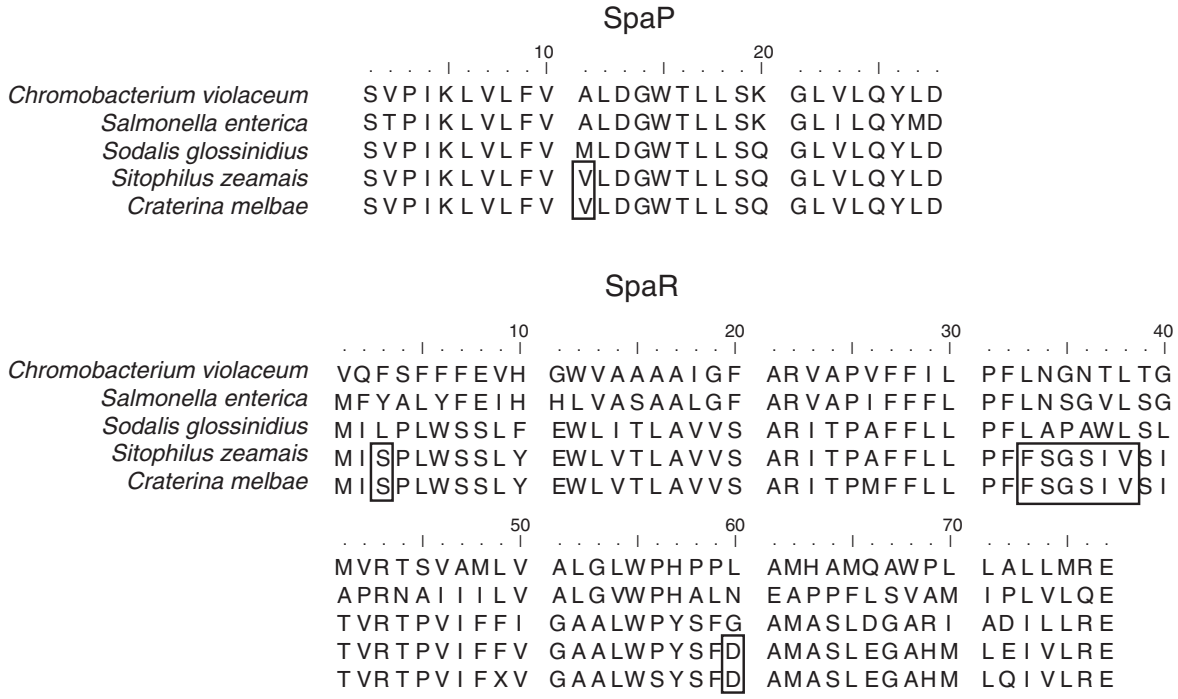


Fig. 2. Aligned regions of SpaP and SpaR genes. The potential synapomorphies are designated by open boxes.

distance method was used based on a nonhomogeneous model implemented in the program PhyloWin (Galtier *et al*., 1996), the symbiont from *Sitophilus rugicollis* clustered at the base of all *Sodalis* sequences (Fig. 1a). Nevertheless, it is important to stress that the *Sitophilus rugicollis* issue is not crucial for solving the *Craterina melbae*-symbiont phylogeny and has not been thoroughly addressed in the present analysis (in fact, it is suspected that the available 16S rRNA gene information is not sufficient to solve the phylogenetic position of all *Sodalis*-related symbionts). The most important finding in this respect is that regardless of the exact position of the *Sitophilus rugicollis* symbiont, the sequence derived from the *Craterina melbae* invariantly clusters as a sister group of the bacteria isolated from the three related *Sitophilus* species (*Sitophilus zeamais, Sitophilus oryzae and Sitophilus granarius*; Fig. 1a). The finding of an independent *Sodalis* lineage in another bloodsucking hippoboscoid species indicates that this bacterial lineage may be more broadly distributed than evidenced by the available symbiont records. While such a possibility might solve the contradiction between the monophyly and low diversity of *Sodalis glossinidius* (by showing that the *Sodalis glossinidius* is not monophyletic but only represents a sample from more broadly distributed bacterium), it raises the question of why these bacteria are so overlooked. One typical difficulty with screening for the symbiotic bacteria is associated with the specificity of 16S rRNA gene primers. It is shown elsewhere (Hypsa & Krizek, 2007) that specifically designed primers can detect a lineage of symbiotic bacteria that are otherwise 'invisible' to regularly used universal primers. This may occur either due to the predominance of other bacterial symbionts in the same sample or due to the mutations within the critical region of the primer sequence. As the symbiotic bacteria are known to accumulate mutations within their genomes, including otherwise conservative regions, the danger of losing priming sites is likely to be higher than in free-living bacteria. To facilitate screening for other *Sodalis* lineages, a specific 16S rRNA gene primer was

designed based on the known *Sodalis* sequences (see 'Materials and methods'). We tested these primers using the *Craterina*-derived symbiont as a positive sample and another S symbiont, *Arsenophonus triatominarum*, as the negative sample, and their specificity was confirmed for *Sodalis* 16S rRNA gene. These markers, together with the Spa-specific primers, might provide a tool allowing for identification of other *Sodalis* lineages and a more complex investigation of the distribution and biology of these interesting symbiotic bacteria.

Acknowledgements

The authors are grateful to Pierre Bize for providing the *Craterina* samples. This work was supported by Grant 206/04/0520 (Grant Agency of the Czech Republic) and the grants LC06073 and MSM 60076605801 (Ministry of Education, Czech Republic).

References

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.

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

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

Dale C, Young SA, Haydon DT & Welburn SC (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *Proc Natl Acad Sci USA* **98**: 1883–1888.

Galtier N & Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. *Proc Natl Acad Sci USA* **92**: 11317–11321.

Galtier N & Gouy M (1998) Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol Biol Evol* **15**: 871–879.

Galtier N, Gouy M & Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12**: 543–548.

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

Geiger A, Ravel S, Frutos R & Cuny G (2005b) *Sodalis glossinidius* (*Enterobacteriaceae*) and vectorial competence of *Glossina palpalis gambiensis* and *Glossina morsitans morsitans* for *Trypanosoma congolense* savannah type. *Curr Microbiol* **51**: 35–40.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acid Symp Ser* **41**: 95–98.

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

Herbeck JT, Degnan PH & Wernegreen JJ (2005) Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the *Enterobacteriales* (*c*-Proteobacteria). *Mol Biol Evol* **22**: 520–532.

Hypsa V & Krizek J (2007) Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). *Microbial Ecol* (in press).

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.

Maudlin I & Ellis DS (1985) Association between intracellular rickettsia-like infections of midgut cells and susceptibility to trypanosome infection in *Glossina* spp. *Z Parasitenkunde* **71**: 683–687.

Moloo SK & Shaw MK (1989) Rickettsial infections of midgut cells are not associated with susceptibility of *Glossina morsitans centralis* to *Trypanosoma congolense* infection. *Acta Trop* **46**: 223–227.

Nardon P, Lefevre C, Delobel B, Charles H & Heddi A (2002) Occurrence of endosymbiosis in *Dryophthoridae* weevils: cytological insights into bacterial symbiotic structures. *Symbiosis* **33**: 227–241.

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

Posada D & Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Swofford DL (1998) *PAUP 4.0 – Phylogenetic Analysis Using Parsimony. Version 4*. Sinauer Associates, Sunderland, MA.

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

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

Paper 7:

Nováková, E., Hypša, V., & Moran, N. A. (2009). *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC microbiology*, *9*(1), 143.

BMC Microbiology

BioMed Central

Research article

Open Access

Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution

Eva Nováková*[1], Václav Hypša[2] and Nancy A Moran[3]

Address: [1]Faculty of Science, University of South Bohemia, Branišovská 31, České BudEjovice 37005, Czech Republic, [2]Faculty of Science, University of South Bohemia and Institute of Parasitology, Biology Centre of ASCR, Branišovská 31, České BudEjovice 37005, Czech Republic and [3]Department of Ecology and Evolutionary Biology, The University of Arizona, 1041 E. Lowell St, Tucson, Arizona 85721-0088, USA

Email: Eva Nováková* - novaeva@paru.cas.cz; Václav Hypša - vacatko@paru.cas.cz; Nancy A Moran - nmoran@email.arizona.edu

* Corresponding author

Published: 20 July 2009

BMC Microbiology 2009, **9**:143 doi:10.1186/1471-2180-9-143

Received: 25 June 2009

Accepted: 20 July 2009

This article is available from: http://www.biomedcentral.com/1471-2180/9/143

Abstract

Background: The genus *Arsenophonus* is a group of symbiotic, mainly insect-associated bacteria with rapidly increasing number of records. It is known from a broad spectrum of hosts and symbiotic relationships varying from parasitic son-killers to coevolving mutualists.

The present study extends the currently known diversity with 34 samples retrieved mainly from hippoboscid (Diptera: Hippoboscidae) and nycteribiid (Diptera: Nycteribiidae) hosts, and investigates phylogenetic relationships within the genus.

Results: The analysis of 110 *Arsenophonus* sequences (incl. *Riesia* and *Phlomobacter*), provides a robust monophyletic clade, characterized by unique molecular synapomorphies. On the other hand, unstable inner topology indicates that complete understanding of *Arsenophonus* evolution cannot be achieved with 16S rDNA. Moreover, taxonomically restricted *Sampling* matrices prove sensitivity of the phylogenetic signal to sampling; in some cases, *Arsenophonus* monophyly is disrupted by other symbiotic bacteria. Two contrasting coevolutionary patterns occur throughout the tree: parallel host-symbiont evolution and the haphazard association of the symbionts with distant hosts. A further conspicuous feature of the topology is the occurrence of monophyletic symbiont lineages associated with monophyletic groups of hosts without a co-speciation pattern. We suggest that part of this incongruence could be caused by methodological artifacts, such as intragenomic variability.

Conclusion: The sample of currently available molecular data presents the genus *Arsenophonus* as one of the richest and most widespread clusters of insect symbiotic bacteria. The analysis of its phylogenetic lineages indicates a complex evolution and apparent ecological versatility with switches between entirely different life styles. Due to these properties, the genus should play an important role in the studies of evolutionary trends in insect intracellular symbionts. However, under the current practice, relying exclusively on 16S rRNA sequences, the phylogenetic analyses are sensitive to various methodological artifacts that may even lead to description of new *Arsenophonus* lineages as independent genera (e.g. *Riesia* and *Phlomobacter*). The resolution of the evolutionary questions encountered within the *Arsenophonus* clade will thus require identification of new molecular markers suitable for the low-level phylogenetics.

Background

The bacterial genus *Arsenophonus* corresponds to a group of insect intracellular symbionts with a long history of investigation. Although many new *Arsenophonus* sequences have been published in the last several years, along with documentation of diverse evolutionary patterns in this group (Figure 1), the first records of these bacteria date to the pre-molecular era. Based on ultrastructural features, several authors described a transovarially transmitted infection associated with son-killing in the parasitoid wasp *Nasonia vitripennis* [1-3]. Later, they were formally assigned to a new genus within the family Enterobacteriaceae with a single species, *Arsenophonus nasoniae* [4]. The same authors proposed a close relationship of *Arsenophonus* to free-living bacteria of the genus *Proteus*. Independently, other microscopic studies revealed morphologically similar symbionts from various tissues of blood-sucking triatomine bugs [5,6]; a decade later these bacteria were determined on molecular grounds to belong to the same clade and were named *Arsenophonus triatominarum* [7]. Interestingly, the next record on symbiotic bacteria closely related to *A. nasoniae* was from a phytopathological study investigating marginal chlorosis of strawberry [8]. Since available sequence data were insufficient for reliable phylogenetic placement, the phloem-inhabiting pathogen was described as a new genus, *Phlomobacter*, with a single species *P. fragariae* [8].

Since these descriptions, the number of *Arsenophonus* records has steadily been increasing, resulting in two important changes in knowledge of *Arsenophonus* evolution and roles in hosts. First, the known host spectrum has been considerably extended with diverse insect groups and even non-insect taxa. So far, *Arsenophonus* has been identified from parasitic wasps, triatomine bugs, psyllids, whiteflies, aphids, ticks, ant lions, hippoboscids, streblids, bees, lice, and two plant species [4,7-23]. Second, these recent studies have revealed an unsuspected diversity of symbiotic types within the genus. This dramatically changes the original perception of *Arsenophonus* as a bionomically homogeneous group of typical secondary ("S-") symbionts undergoing frequent horizontal transfers among phylogenetically distant hosts. For example, recent findings indicate that some insect groups harbor monophyletic clusters of *Arsenophonus*, possibly playing a role of typical primary ("P-") symbionts. These groups were reported from the dipteran families Hippoboscidae and Streblidae [20] and most recently from several lice species [18,24,25]. Such a close phylogenetic relationship of different types of symbiotic bacteria is not entirely unique among insect symbionts. With the increasing amount of knowledge on the heterogeneity and evolutionary dynamics of symbiotic associations, it is becoming clear that no distinct boundaries separate the P- and S-symbionts. Thus, in their strict meaning, the terms have recently become insufficient, especially for more complex situations, such as studies exploring bacterial diversity within a single host species [14,17]. Furthermore, these terms have been shown not to reflect phylogenetic position; remarkable versatility of symbiotic associations can be observed

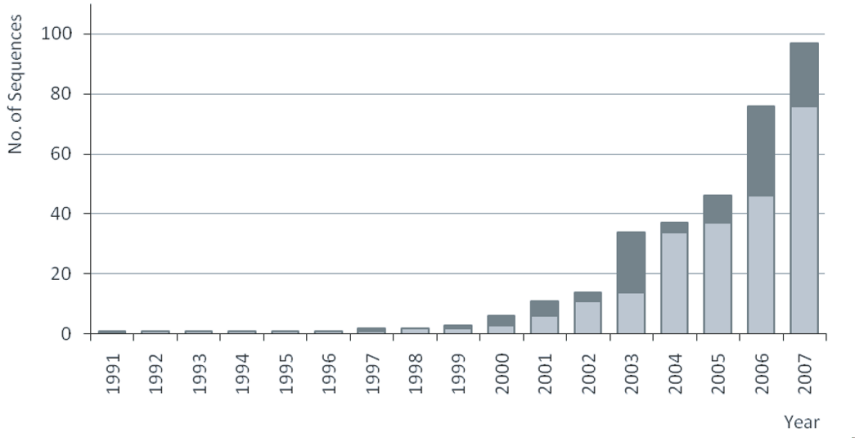


Figure 1

An increase of records on *Arsenophonus* bacteria from various insect groups. The bars show cumulative numbers of sequences deposited into GenBank; dark tops represent new records added in the given year. The sequences are identified with the following accession numbers: 1991 – M90801; 1997 – U91786; 2000 – AF263561, AF263562, AF286129, AB038366; 2001 – AF400474, AF400480, AF400481, AF400478, AY057392; 2002 – AY136168, AY136153, AY136142; 2003 – AY265341–AY265348, Y264663–AY264673, AY264677; 2004 – AY587141, AY587142, AY587140; 2005 – DQ068928, DQ314770–DQ314774, DQ314777, DQ314768, DQ115536; 2006 – DQ538372–DQ538379, DQ508171–DQ508186, DQ517447, DQ508193, DQ837612, DQ837613; 2007 – EU039464, EU043378, EF110573, EF110574, DQ076660, DQ076659, EF110572, EF647590, AB263104.

in the Gammaproteobacteria overall, as well as within the individual clusters, such as *Arsenophonus* or *Sodalis* [16,26].

The genus *Arsenophonus* is striking in the diversity of symbiont types represented. Apart from many lineages with typical S-symbiont features, this genus has given rise to several clusters of P-symbionts [18,20,24]. Unfortunately, this heterogeneity introduced an annoying degree of phylogenetic instability and nomenclatory confusion. Because P-symbionts show accelerated evolutionary rates, they form long branches in phylogenies, leading to unstable patterns of clustering as observed for P-symbionts within Enterobacteriaceae [27]. The same behavior can be seen in the louse-specific clade of *Arsenophonus*, which are consequently originally described as a new bacterial genus *Riesia* [25]. In addition, the *Arsenophonus* cluster is the only monophyletic group of symbiotic bacteria currently known to possess at least four highly different phenotypes, including son-killing [4], phytopathogenicity [8], obligate association with bacteriocytes in the host [18,20,24], and apparently non-specific horizontally transmitted bacteria that are possibly mutualistic [15]. These characteristics indicate that the genus *Arsenophonus* represents an important and widespread lineage of symbiotic bacteria that serves as a valuable model for examining molecular evolution of bacteria-arthropod associations.

In this study, we add 34 new records on symbionts to the known spectrum of *Arsenophonus* lineages. We explore and summarize the current picture of *Arsenophonus* evolution by analyzing all sequences available for this clade. To investigate the phylogenetic position, stability and evolutionary trends of the *Arsenophonus* cluster, we complete the sample with related symbionts and free-living bacteria. Finally, we explore molecular characteristics and informative value of the 16S rRNA gene as the most frequently used phylogenetic marker.

Results

Sequences and alignments

From 15 insect taxa, we obtained 34 sequences of 16S rDNA that exhibited a high degree of similarity to sequences from the bacterial genus *Arsenophonus* when identified by BLAST. The length of the PCR-amplified fragments varied from 632 to 1198 bp, with the guanine-cytosine (GC) content ranging from 46.22 to 54.84% (Figure 2, bars). For three specimens of the hippoboscid *Ornithomya avicularia*, two different sequences were obtained from each single individual. After combining with all *Arsenophonus* 16S rDNA sequences currently available in the GenBank, and several additional free-living and symbiotic bacteria, the dataset produced a 1222 bp long *Basic matrix*. The alignment has a mosaic structure, discussed below. Within the set, a large group of sequences show a high degree of similarity (0.1–7.3% divergence) and exhibit GC content and sequence length similar to those found in free-living enterobacteria. The set also includes several sequences with modifications typical for many proteobacterial symbionts, particularly the presence of long insertions within the variable regions and decreased GC content. Sequence distances among these taxa range up to 17.8%.

Phylogeny

All phylogenetic analyses of the *Basic matrix* yielded a monophyletic *Arsenophonus* clade (Figure 2). The new 34 sequences (Figure 2, arrows), identified by BLAST as putative members or relatives of the genus *Arsenophonus*, always clustered within the *Arsenophonus* clade. Their precise position was only partially correlated with host taxon. Some of the *Arsenophonus* sequences from hippoboscoid hosts clustered within monophyletic host-specific groups (Figure 2, printed in red) while others were scattered across the tree as isolated lineages (Figure 2, printed in dark orange). Two distinct sequences were determined from each individual specimen of *O. avicularia*; these clustered at distant positions within the tree (Figure 2, numbers with asterisks).

The most typical lineages display short-branches with low divergence and unstable positions within the *Arsenophonus* clade (Figure 2, printed in dark orange). At the opposite extreme are well supported host-specific clusters exhibiting long branches, such as the louse symbiont *Riesia* or the symbionts described from several streblid species. An intermediate situation is found in putatively host-specific but less robust clusters, such as the *Arsenophonus* lineages from triatomine bugs, some hippoboscoids or homopterans (Figure 2). In an analogy to previously analyzed symbiotic bacteria [e.g. [28,29]], the phylogenetic properties of the sequences were also reflected in their GC contents. In the short-branched taxa, the GC content of the 16S rRNA sequence varies from 51.72 to 54.84%, the values typical for S-symbionts and free-living bacteria [30]. In contrast, the 16S rRNA sequences with low GC content, varying between 46.22 and 51.93%, were found in the long-branched taxa clustering within the host-specific monophyletic lineages (e.g. the symbionts from *Ornithomyia*, *Lipoptena*, *Trichobius*, and the *Riesia* clade).

Considerable loss of phylogenetic information was observed in the *Conservative matrix*. In this case, the relationships among individual *Arsenophonus* lineages were highly unstable, resulting in large polytomies of many short-branched taxa within the consensus trees (see Additional file 1). Also the relationships among the long-branched lineages, although resolved, differ sharply from those derived from the *Basic matrix* data, and the genus

Figure 2
Phylogenetic tree derived from the *Basic matrix* (1222 positions) under ML criterion. Names of the taxa clustering within the *Arsenophonus* clade are printed in colour: red for the long-branched taxa, dark orange for the short-branched taxa. The arrows point to the new sequences obtained in our study. Different types of sequences determined from the specimens of *O. avicularia* are designated by the numbers with asterisks. The type species *A. nasoniae* is designated by the orange asterisk. Solid circles on branches label the clusters strictly concordant with the host phylogenies. Open circles designate host-specific lineages without coevolutionary signal. Solid vertical lines indicate reciprocally monophyletic groups of symbionts and hosts. Dashed lines show paraphyletic symbiont clades restricted to monophyletic host groups. Names in the brackets indicate host taxa. "Symb-" in the taxon designation stands for "Symbiotic bacteria of". Bars represent GC content of each taxa. Complete information on the sequences is provided in the Additional file 5.

Proteus was not positioned as the closest relative of *Arsenophonus*. Thus, the information contained in the *Conservative matrix* (restricted to one fourth of *Basic* dataset, i.e., 284 bp) is insufficient for reliable phylogenetic placement of closely related taxa.

The analyses of taxonomically restricted *Sampling* matrices confirmed the expected dependence of the phylogenetic conclusions on the taxon sampling (examples of topologies obtained are provided in Figures 3, 4 and Additional file 2). The highest degree of susceptibility was observed with MP, particularly under Tv:Ts ratio set to 1. The most fundamental distortion occurred with the matrix *Sampling3*, where one lineage (composed of *Buchnera*, *Wigglesworthia*, *Blochmannia*, and S-symbiont from *Trioza magnoliae*) clustered either as a sister group of *Riesia* clade or together with *Sodalis*. Thus, the consensus tree did not preserve the monophyly of an *Arsenophonus* clade (Figure 3).

The calculation of divergence times yielded substantially different results depending on the choice of calibration points. Use of the Riesia diversification as a reference point suggested a recent origin of the triatomine-associated Arsenophonus branch; the median value of the estimate distribution was 2.6 mya. In contrast, the calibration by Escherichia-Salmonella returned considerably higher dates with the median at 24.5 mya.

Discussion

Phylogenetic patterns and the stability of the information

Phylogenetic relationships of the *Arsenophonus* symbionts display a remarkably complex arrangement of various types of symbiosis and evolutionary patterns. Moreover, a comparison of the branch ordering within each of these subclusters to the host phylogeny indicates a cospeciation process within several lineages (discussed below). From the phylogenetic perspective, no clearcut boundary divides the set of *Arsenophonus* sequences into the ecologically distinct types. The position of the long-branched subclusters within the topology is not stable. Under the MP criterion with transition rate 1:3 and under the ML criterion they form a unique monophyletic cluster (Figure 5A), while in other analyses the individual host-specific subclusters were scattered among the short-branched lineages (Figure 5B, Figure 6).

The low resolution and instability of the trees inferred from the *Conservative matrix* suggest that a substantial part of the phylogenetic information is located within the "ambiguously" aligned regions that were removed by the GBlocks procedure. This fact is particularly important when considering the frequent occurrence of insertions/deletions within the sequences (see Additional file 3). This may lead to deletion of these critical fragments in many phylogenetic analyses. Interestingly, the monophyletic nature of *Arsenophonus* was preserved even in this highly *Conservative matrix*. This indicates that within the complete data set, the phylogenetic information underlying the *Arsenophonus* monophyly is sufficiently strong and is contained in the conservative regions of the sequences. In accordance with this presumption, several molecular synapomorphies can be identified in the *Basic* and *Conservative* matrices. The most pronounced is the motif GTC/GTT located in positions 481–483 and 159–161 of *Basic matrix* and *Conservative matrix*, respectively.

Relevance of the sampling

To test an effect of sampling on the phylogenetic inference within *Arsenophonus*, we examined five *Sampling* matrices with different taxa compositions (see the section *Methods*). In addition to the MP, ML, and Bayesian analyses, we performed an ML calculation under the nonhomogeneous model of the substitutions, designated as T92 [31,32]. This model was previously used to test the monophyly/polyphyly of the P-symbiotic lineages and brought the first serious evidence for a possible independent origin of major P-symbiotic taxa [27]. We were not able to apply the same approach to the *Basic* and *Conservative* matrices since the program Phylowin failed to process these large datasets under the ML criterion. The analyses of several taxonomically restricted *Sampling* matrices proved the sensitivity of phylogenetic signal to the sampling. In the most extreme case, shown in Figure 3A, even the monophyly of the *Arsenophonus* clade was disrupted by other lineages of symbiotic bacteria. Considering the results of the extensive analysis of the *Basic matrix*, this arrangement is clearly a methodological artifact. Since both *Riesia* and the P-symbiont lineage are long-branched taxa with rapid evolution of 16S rDNA, their affinity is very likely caused by Long Branch Attraction (LBA; for review see [33]) within the taxonomically compromised matrix. It is symptomatic that this topology was inferred by MP, the method known to be particularly prone to the LBA. To further test this distortion, one of the long-branched taxa was removed from the data set (matrix *Sampling4*). This approach restored the *Arsenophonus* monophyly and confirmed the effect of LBA phenomenon (see Additional file 2).

The aim of these taxonomically restricted analyses was to "simulate" phylogenetic placement of newly determined symbionts. In such casual studies, the symbiotic lineages are rarely represented by all available sequences in the way we composed the *Basic matrix*. Rather, each symbiotic lineage is represented by few randomly selected sequences. Under such circumstances, incorrect topologies (e.g. the *Sampling5*-derived topology on the Figure 4) can be obtained due to various methodological artifacts. This situation can be illustrated by empirical data: at least in two

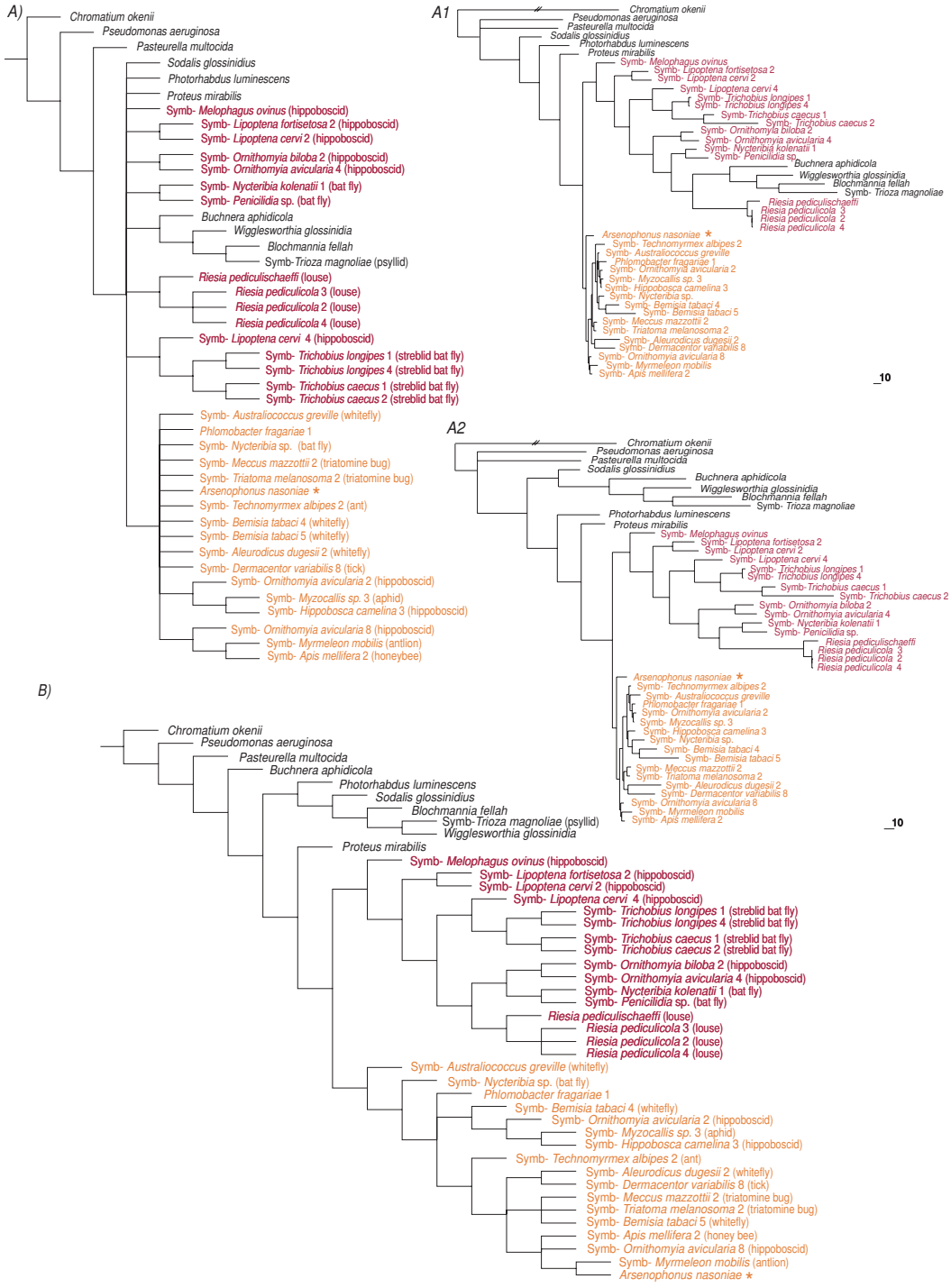


Figure 3
Topologies derived from *Sampling3* matrix (851 positions). A) consensus of the trees and two tree examples A1 and A2, obtained under the MP criterion with Tv/Ts ratio set to 1:1 B) consensus of the trees obtained under the MP criterion with Tv/Ts ratio set to 1:3. The type species *A. nasoniae* is designated by the orange asterisk.

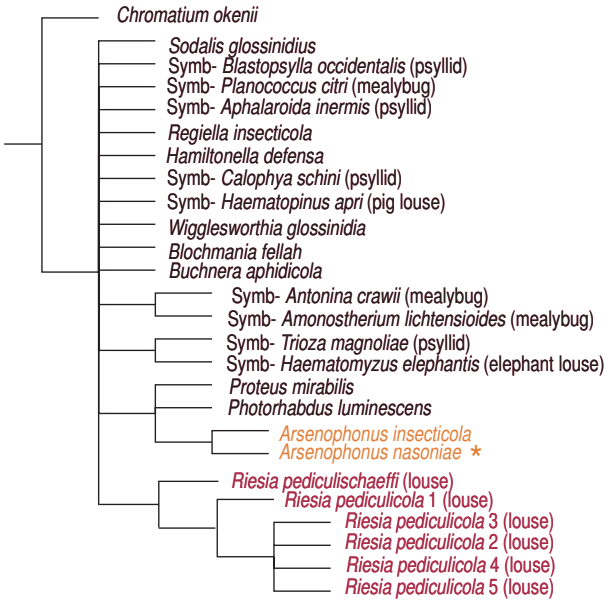


Figure 4
Tree consensus derived from *Sampling5* (936 positions) matrix under the MP criterion. Transversion/transition ratio was set to 1:1. The type species *A. nasoniae* is designated by the orange asterisk.

studies, the louse-associated lineage of *Arsenophonus* was not recognized as a member of the *Arsenophonus* clade [25,34]. Consequently, when more recent studies, based on better sampling, proved the position of *Riesia* within the *Arsenophonus* cluster [18,24] the genus *Arsenophonus* became paraphyletic (see the section *Conclusion* for more details).

Interestingly, topologies inferred by likelihood analyses using the T92 evolution model [31] were influenced neither by the compromised sampling nor by the removal of unreliably aligned regions.

Cophylogeny vs. horizontal transfers: possible sources of phylogenetic incongruence

The phylogenetic tree of all *Arsenophonus* sequences exhibits both patterns, the parallel evolution of symbionts and their hosts and the haphazard association of symbionts from different host taxa. Coincidentally, both arrangements can be demonstrated on the newly sequenced symbionts from various hippoboscoid species. Some of hippoboscoid-associated *Arsenophonus* show possible host specificity; in a few analyses they cluster within several monophyletic short-branched groups. Since relationships among the short-branched taxa are generally not well resolved, these lineages are scattered throughout the whole topology (Figure 2). In contrast, relationships within the long-branched clusters of hippoboscoid-associated taxa are in agreement with the host phylogeny (the *Arsenophonus* clusters strictly reflecting the host phylogeny are designated by solid circle in the Figure 2). Interestingly, a coevolutionary pattern was also identified for streblids of the genus *Trichobius* and their symbionts. In the original study published by Trowbridge et al. [20], the distribution of *Trichobius* symbionts was apparently not consistent with the host phylogeny. Our analysis in a broad context indicates that this discrepancy might have been caused by different bacterial sampling and particularly by aberrant behavior of the sequence from *Trichobius yunkeri* [GenBank: DQ314776]. This sequence is likely to be an artificial chimerical product of at least two distant lineages; according to our BLAST tests it shares 100% identity with S-symbiont of *Psylla pyricola* [GenBank: AF286125] along a 1119 bp long region. Removal of this sequence from the dataset restored a complete phylogenetic congruence between *Trichobius*, based on the phylogeny of this genus published by Dittmar et al. [35], and its symbionts. This finding exemplifies the danger of chimeric sequences in studies of symbiotic bacteria, obtained by the PCR on the sample containing DNA mixture from several bacteria. The presence of several symbiotic lineages within a single host is well known [e.g. [14,36-38]]. In this study, we demonstrate a possible such case in *O. avicularia*. From three individuals of this species we obtained pairs of different sequences branching at two distant positions (labelled by the numbers 1* to 3* in Figure 2). The identical clustering seen in all three pairs within the tree shows that they are not chimeric products but represent two different sequences.

While the identity between symbiont relationships and the host phylogeny is apparently a consequence of host-symbiont cophylogeny, the interpretation of the randomly scattered symbionts is less obvious. Usually, such an arrangement is explained as result of transient infections and frequent horizontal transfers among distant host taxa. This is typical, for example, of the *Wolbachia* symbionts in wide range of insect species [39]. Generally, the capability to undergo inter-host transfers is assumed for several symbiotic lineages and has even been demonstrated under experimental conditions [40,41]. Since the *Arsenophonus* cluster contains bacteria from phylogenetically distant insect taxa and also bacteria isolated from plants, it is clear that horizontal transfers and/or multiple establishments of the symbiosis have occurred. However, part of the incongruence could be caused by methodological artifacts. A conspicuous feature of the *Arsenophonus* topology is the occurrence of monophyletic symbiont lineages associated with monophyletic groups of insect host but without a co-speciation pattern. Although our study cannot present an exhaustive explanation of such a picture, we want to point out two factors that might in theory

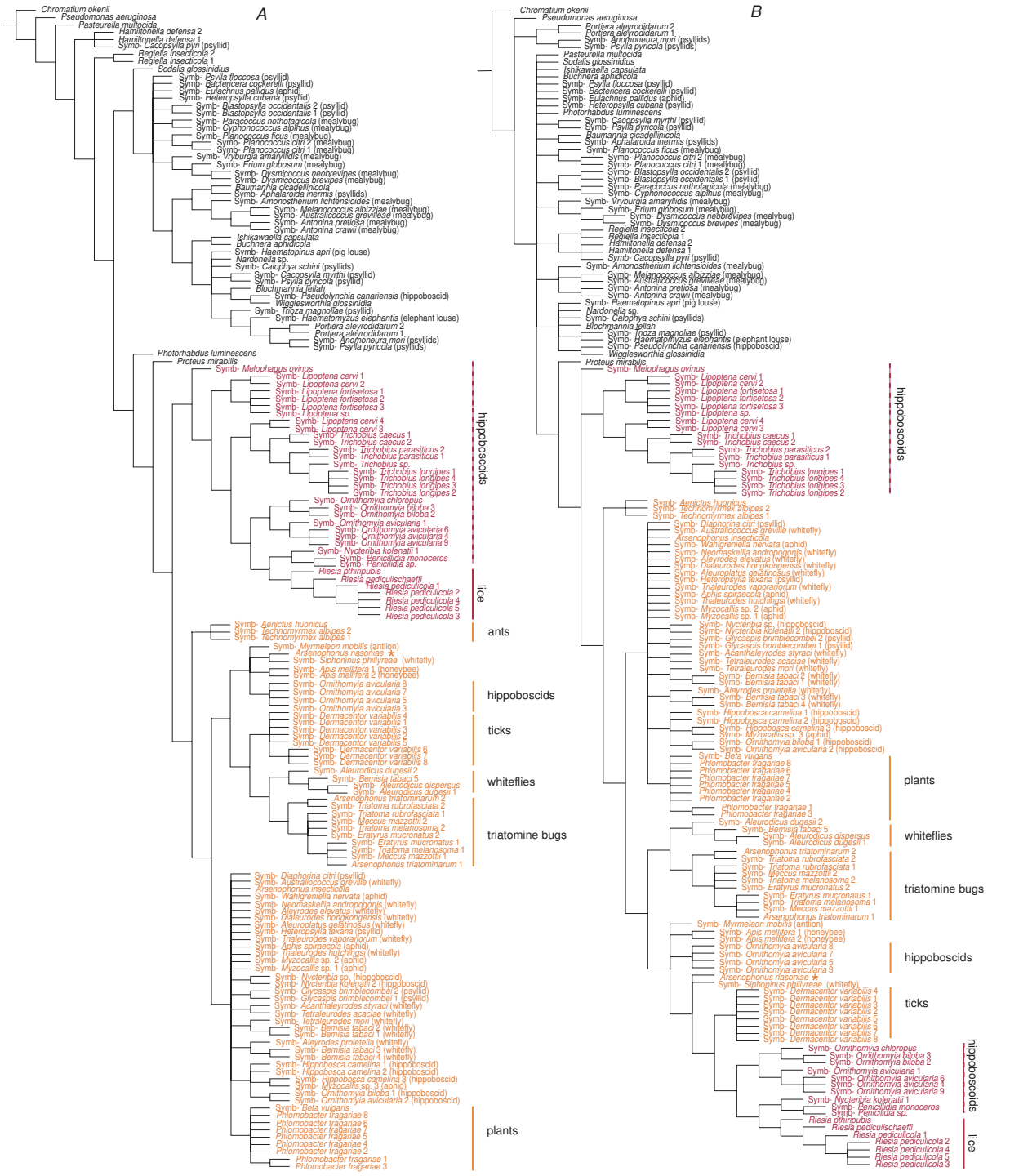


Figure 5
Topologies derived from the *Basic matrix* (1222 positions). A) consensus of the trees obtained under the MP criterion with transversion/transition ratio set to 1:3 and the ML criterion; B) consensus of the MP trees obtained with the transversion/transition ratio 1:1. The type species *A. nasoniae* is designated by the orange asterisk.

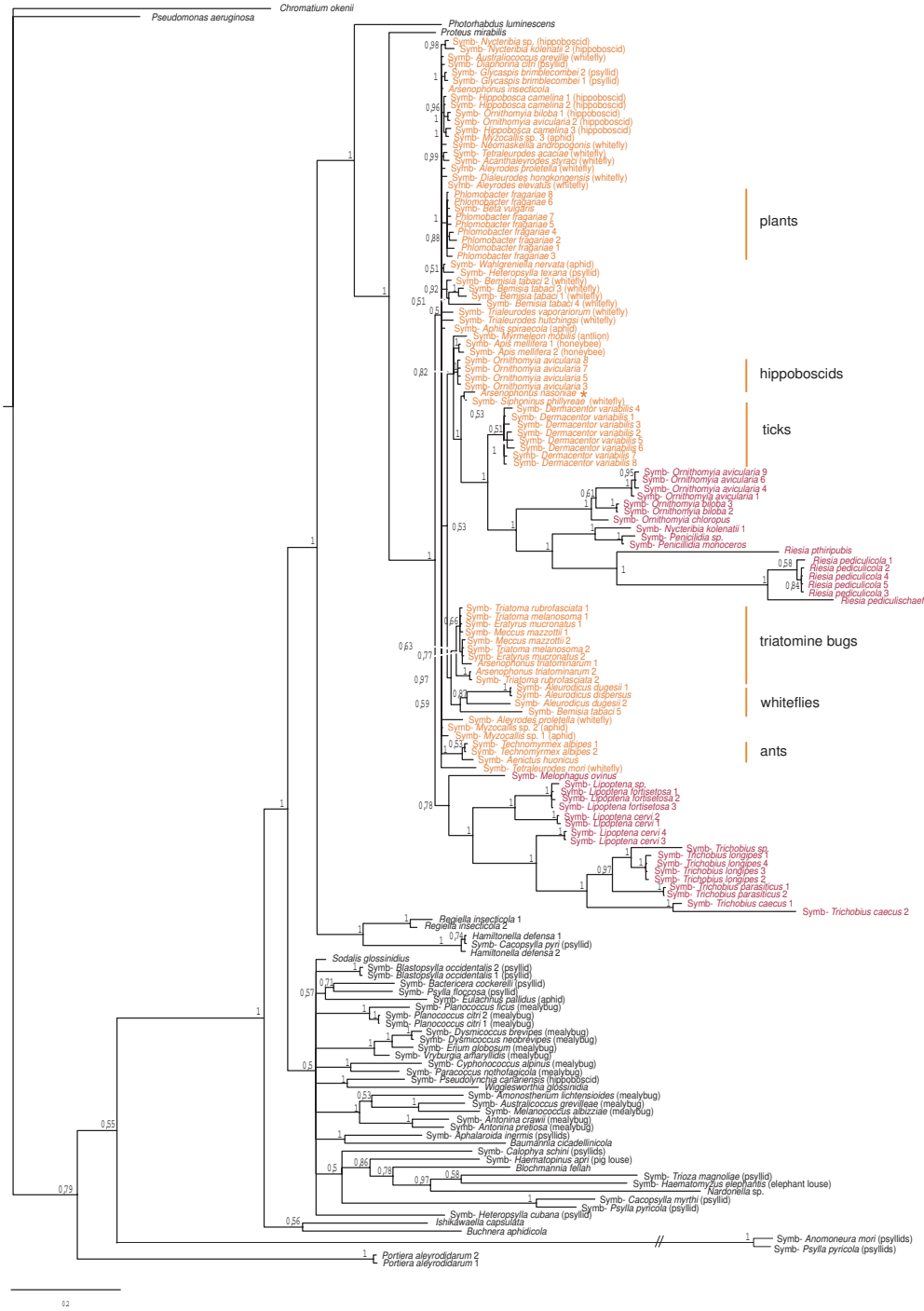


Figure 6

Phylogenetic tree derived from *Basic matrix* (1222 positions) using Bayesian analysis. Names of the taxa clustering within the *Arsenophonus* clade are printed in colour: red for the long-branched taxa, dark orange for the short-branched taxa. Names in the brackets designate the host family. Numbers represent Bayesian posterior probability for each node. The type species *A. nasoniae* is designated by the orange asterisk.

take part in shaping the relationships among *Arsenophonus* sequences, lateral gene transfer (LGT) and intragenomic heterogeneity. Both have previously been determined as causes of phylogenetic distortions and should be considered in coevolutionary studies at a low phylogenetic level.

Incongruence due to LGT and intragenomic heterogeneity

An apparently "mosaic" structure of the *Arsenophonus* alignment (for example see Additional file 4) raises the question of whether various regions of this sequence could have undergone different evolutionary histories. Recombination of 16S rDNA genes were previously identified in some other bacteria [42-44]. In actinomycetes, the occurrence of short rDNA segments with high number of non-random variations was attributed to the lateral transfer as the most parsimonious explanation [45]. Later, Gogarten et al. [46] suggested that, analogously to an entire bacterial genome, 16S rDNA possesses a mosaic character originated by LGT, respectively by transfer of gene subunits.

As bacterial genomes often carry more than one rRNA operon, intragenomic heterogeneity of the rDNA copies is occasionally found to blur the phylogenetic picture [47-50]. Although there is no direct information on the number of rRNA gene copies in *Arsenophonus* genomes, Stewart and Cavanaugh [51] showed bacterial genomes to encode in average five rRNA operons. The most closely related bacterium of which the complete genome has recently been sequenced, *Proteus mirabilis*, carries seven copies [GenBank: AM942759]. *Arsenophonus*-focused studies indicate that two different forms of the rRNA operon are present in its genome, as is typical for Enterobacteraceae [23,52]. Furthermore, Šorfová et al. [23] suggest that the variability among individual copies may cause the incongruence observed between triatomines and their *Arsenophonus* lineages. They point out that this process could, in principle, explain an otherwise problematic observation: in some hosts, such as triatomines or some homopterans, the hosts and the *Arsenophonus* bacteria create reciprocally monophyletic clusters but do not show any cospeciation pattern. In the symbionts of grain weevils, divergence between rRNA sequences within a genome was shown sometimes to exceed divergence of orthologous copies from symbionts from different hosts; this unusual situation was hypothesized to reflect loss of recombinational repair mechanisms from these symbiont genomes [53].

Estimates of the divergence time

With the present incomplete knowledge of the *Arsenophonus* genome, it is difficult to assess whether and how deeply rRNA heterogeneity affects phylogenetic reconstruction. Trying to find alternative solution, Šorfová et al. [23] attempted to use the estimation of divergence times as a guide for deciding between different coevolutionary scenarios. They used the *Escherichia-Salmonella* divergence [54,55] as a calibration point for calculating the divergence time among various *Arsenophonus* lineages from triatomine bugs. Applying the Multdiv method [56], they placed the ancestor of triatomine-associated symbionts into a broad range of approx. 15 – 40 mya and concluded that this estimate is compatible or even exceeds the age estimates available for the tribe triatomine (according to Gaunt and Miles [57]). Here, we took advantage of a new age-estimate for closely related bacteria, namely the louse-associated symbionts of the genus *Riesia* [18]. Comparing the estimates based on the two calibration methods (*Escherichia-Salmonella* and *Riesia*), we found that due to the variability of evolutionary rates among the lineages, the results may differ by an order of magnitude. Such marked variance among different bacterial lineages (including different symbiotic bacteria from the same host species) was previously reported for many bacterial groups [29,30,37,39,58-63]. Most recently, Allen et al. [64] reported an extremely high evolutionary rate for the young symbiotic lineage *Riesia*, and suggested that the evolutionary tempo changes with the age of the symbiotic lineage. We therefore conclude that this method cannot be directly used to assess the effect of intragenomic heterogeneity on our reconstruction of *Arsenophonus* relationships.

Conclusion

With more than one hundred records, the genus *Arsenophonus* represents one of the richest and most widespread clusters of insect symbiotic bacteria. Considering its broad host spectrum and apparent ecological versatility, *Arsenophonus* should play an important role in studies of evolutionary trends in insect intracellular symbionts. Due to this fact, *Arsenophonus* is likely to attract a growing attention, and the number of the records may rapidly be increasing during the next years. For example, 7 new sequences were deposited into the GenBank since the completion of this study [[65], and unpublished record FJ388523]. However, since these new *Arsenophonus* records originated in screening rather than phylogenetic study, they are only represented by short DNA fragments (approx. 500 bp). Preliminary analyses of these fragments together with our complete datasets confirmed a limited informative value of such short sequences and they were not included into the more exhaustive phylogenetic procedures.

The analysis of 110 available sequences of *Arsenophonus* 16S rDNA from 54 host taxa revealed several interesting evolutionary patterns. In particular, this clade includes at least two transitions from S-symbiont, with ability to invade new host lineages, to P-symbiont, showing obligate relationship to hosts and a strict pattern of maternal

transmission. Thus, it is a promising system for exploring the genomic and biological changes that accompany the shift from facultative to obligate symbiont. *Arsenophonus* is also one of the few groups of insect symbionts for which strains have been grown in pure culture [4,7,16], a feature that further enhances its potential as a model for symbiont research.

Our results also indicate that a complete understanding of the *Arsenophonus* phylogeny cannot be achieved with 16S rDNA genes alone. A similar situation is, for example, found in another large symbiotic group, the genus *Wolbachia*, where other genes are often used as alternative sources of phylogenetic information [66,67]. Identification of suitable low-level-phylogeny marker(s) is thus one of the most crucial steps in the further research on *Arsenophonus* evolution. The sequencing of the complete *Arsenophonus* genome, which is currently under the process http://genomesonline.org/gold.cgi?want=Bacterial+Ongoing+Genomes&pubsort=Domain, will provide a valuable background for such enterprise.

Based on the presented analyses, we also want to point out that the genus *Arsenophonus* is currently paraphyletic due to the two lineages described as separate genera *Riesia* and *Phlomobacter* but clustering within the *Arsenophonus* group (e.g. Figure 2). Two procedures can, in principle, solve this undesirable situation, splitting of the *Arsenophonus* cluster into several separate genera or classification of all its members within the genus *Arsenophonus*. Taking into account the phylogenetic arrangement of the individual lineages, the first approach would inevitably lead to establishment of many genera with low sequence divergences and very similar biology. The second option has been previously mentioned in respect to the genus *Phlomobacter* [68], and we consider this approach (i.e. reclassification of all members of the *Arsenophonus* clade within a single genus) a more appropriate solution of the current situation within the *Arsenophonus* clade.

Methods

Samples

The host species used in this study were acquired from several sources. All of the nycteribiid samples were obtained from Radek Lučan. Most of the hippoboscids were provided by Jan Votýpka. Ant species were collected by Milan Janda in Papua New Guinea. All other samples are from the authors' collection. List of the sequences included in the *Basic matrix* is provided in the Additional file 5.

DNA extraction, PCR and sequencing

The total genomic DNA was extracted from individual samples using DNEasy Tissue Kit (QIAGEN; Hilden, Germany). Primers F40 and R1060 designed to amplify approx. 1020 bp of 16S rDNA, particularly within Enterobacteriaceae [34], were used for all samples. PCR was performed under standard conditions using HotStart Taq polymerase (HotStarTaqi DNA Polymerase, Qiagen). The PCR products were analyzed by gel electrophoresis and cloned into pGEM-T Easy System 1 vector (Promega). Inserts from selected colonies were amplified using T7 and SP6 primers and sequenced in both directions, with the exception of 3 fragments sequenced in one direction only (sequences from *Aenictus huonicus* and *Myzocalis sp.*). DNA sequencing was performed on automated sequencer model 310 ABI PRISM (PE-Biosystems, Foster City, California, USA) using the BigDye DNA sequencing kit (PE-Biosystems). For each sample, five to ten colonies were screened on average. The contig construction and sequence editing was done in the SeqMan program from the DNASTAR platform (Dnastar, Inc. 1999). Identification of the sequences was done using BLAST, NCBI http://www.ncbi.nlm.nih.gov/blast/Blast.cgi.

Alignments

To analyze thoroughly the behavior of *Arsenophonus* 16S rDNA and assess its usefulness as a phylogenetic marker, we prepared several matrices and performed an array of phylogenetic analyses on each of them.

The *Basic matrix* was composed of the 34 new sequences, all *Arsenophonus* sequences available in the GenBank and additional 45 sequences of various P-symbionts, S-symbiont and 5 free living bacteria (see Additional file 5). To show the impact of random or restricted sampling on the resulting topology, five different matrices labelled *Sampling*$_i$ (i.e. *Sampling1*, *Sampling2*, etc.) were prepared from *Basic matrix* by removing various taxa and including additional/alternative outgroups. The matrices *Sampling1* to *Sampling4* were composed of various numbers of non-*Arsenophonus* symbiotic taxa (ranging from 3 to 35), three sequences of free-living bacteria, and an arbitrarily selected set of all *Arsenophonus* lineages. Matrix designated as *Sampling5* was restricted to a lower number of taxa, including 5 ingroup sequences and alternative lineages of symbiotic and free-living bacteria.

All matrices were aligned in the server-based program MAFFT http://align.bmr.kyushu-u.ac.jp/mafft/online/server/, using the E-INS-i algorithm with default parameters. The program BioEdit [69] was used to manually correct the resulting matrices and to calculate the GC content of the sequences.

To test an effect of unreliably aligned regions on the phylogenetic analysis, we further prepared the *Conservative matrix*, by removing variable regions from the *Basic matrix*. For this procedure, we used the program Gblocks [70] available as server-based application on the web page

http://molevol.cmima.csic.es/castresana/Gblocks_server.html.

Finally, the *Clock matrix*, composed of 12 bacterial sequence (see Additional file 5), was designed to calculate time of divergence for several nodes within the *Arsenophonus* topology.

Phylogenetic analyses

The matrices were analyzed using maximum parsimony (MP), maximum likelihood (ML) and Bayesian probability. For analyses, we used the following programs and procedures. *The GTR+Γ+inv* model of molecular evolution was determined as best fitting by the program Modeltest [71] and was used in all ML-based analyses. MP analysis was carried out in TNT program [72] using the *Traditional search* option, with 100 replicates of heuristic search, under the assumptions of Ts/Tv ratio 1 and 3. ML analysis was done in the Phyml program [73] with model parameters estimated from the data. Bayesian analysis was performed in Mr. Bayes ver. 3.1.2. with following parameter settings: nst = 6, rates = invgamma, ngen = 3000000, samplefreq = 100, and printfreq = 100. The program Phylowin [74] was employed for the ML analysis under the nonhomogeneous model of substitution [31].

A calculation of divergence time was performed in the program Beast [75] which implements MCMC procedure to sample target distribution of the posterior probabilities. The gamma distribution coupled with the *GTR+inv-gamma* model was approximated by 6 categories of substitution rates. Relaxed molecular clock (uncorrelated lognormal option) was applied to model the rates along the lineages. To obtain a time-framework for the tree, we used the estimate on louse divergence (approximately 5.6 mya [18]). Since the resulting estimate was considerably lower that that reported previously with *Escherichia-Salmonella* calibration [23], we prepared an additional matrix and used the *Escherichia-Salmonella* split [54,55] as an alternative calibration; taxa included according to Šorfová et al. [23]. All analyses were performed in three independent runs, each taking 5 million generations.

Authors' contributions

EN obtained the sequence data, compiled alignments and participated in the study design, phylogenetic inference, interpretation of the results, and preparation of the manuscript. VH conceived of the study and participated in conduction of the phylogenetic inference. Both, VH and NAM participated in the study design, evolutionary interpretation of the results and preparation of the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Consensus tree derived from the* Conservative matrix *(284 positions) under MP criterion. *Transversion/transition ratio was set to 1:3. Names of the taxa clustering within the* Arsenophonus *clade derived from* Basic matrix *are printed in colour: red for the long-branched taxa, dark orange for the short-branched taxa.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2180-9-143-S1.eps]

Additional file 2

Tree consensus derived from* Sampling4 *(1107 positions) matrix under the MP criterion. *Transversion/transition ratio was set to 1:1. The type species* A. nasoniae *is designated by the orange asterisk.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2180-9-143-S2.eps]

Additional file 3

Insertions within the sequences of P-like symbionts.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2180-9-143-S3.eps]

Additional file 4

The 41 bp long motif inconsistently distributed among distinct bacterial taxa. *Position of the sequence in alignment and 16S rDNA secondary structure is indicated by the arrows. Following records are not included in the Additional file 1:* Sitophilus rugicollis *[GenBank: AY126639]*, Drosophila paulistorum *[GenBank: U20279, U20278]*, Polyrhachis foreli *[GenBank: AY336986]*, Haematopinus eurysternus *[GenBank: DQ076661]*.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2180-9-143-S4.eps]

Additional file 5

***List of sequences included in* Basic matrix.** *Dashed line separates members of the* Arsenophonus *clade from the outgroup taxa. Sequences included into the* Clock matrix *are underlined.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2180-9-143-S5.doc]

Acknowledgements

This work was supported by Ministry of Education, Czech Republic (grants LC06073 and MSM 60076605801), the Grant Agency of Academy of Sciences of the Czech Republic (Grant IAA601410708) and a National Science Foundation grant (0626716) to N.A. Moran (University of Arizona). We thank all of our collaborators for providing the samples.

References

1. Huger AM, Skinner SW, Werren JH: **Bacterial infections associated with the son-killer trait in the parasitoid wasp, *Nasonia* (= *Mormoniella*) *vitripennis* (Hymenoptera, Pteromalidae).** *J Invertebr Pathol* 1985, **46:**272-280.

2. Skinner SW: **Son-killer – A 3rd extrachromosomal factor affecting the sex-ratio in the parasitoid wasp, *Nasonia* (= *Mormoniella*) *vitripennis*.** *Nasonia* 1985, **109:**745-759.
3. Werren JH, Skinner SW, Huger AM: **Male-killing bacteria in a parasitic wasp.** *Science* 1986, **231:**990-992.
4. Gherna RL, Werren JH, Weisburg W, Cote R, Woese CR, Mandelco L, Brenner DJ: ***Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*.** *Int J Syst Bacteriol* 1991, **41:**563-565.
5. Hypša V: **Endocytobionts of *Triatoma infestans* : distribution and transmission.** *J Invertebr Pathol* 1993, **61:**32-38.
6. Louis C, Drif L, Vago C: **Mise en évidence et étude ultrastructurale de procaryotes de type rickettsien dans les glandes salivaires des Triatomidae (Heteroptera) = Evidence and ultrastructural study of *Rickettsia*-like prokaryotes in salivary glands of Triatomidae (Heteroptera).** *Ann Soc Entomol Fr* 1986, **22:**153-162.
7. Hypša V, Dale C: **In vitro culture and phylogenetic analysis of "Candidatus *Arsenophonus triatominarum*, " an intracellular bacterium from the triatomine bug, *Triatoma infestans*.** *Int J Syst Bacteriol* 1997, **47:**1140-1144.
8. Zreik L, Bove JM, Garnier M: **Phylogenetic characterization of the bacterium-like organism associated with marginal chlorosis of strawberry and proposition of a Candidatus taxon for the organism, 'Candidatus *Phlomobacter fragariae* '.** *Int J Syst Bacteriol* 1998, **48:**257-261.
9. Spaulding AW, von Dohlen CD: **Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA.** *Insect Mol Biol* 2001, **10:**57-67.
10. Subandiyah S, Nikoh N, Tsuyumu S, Somowiyarjo S, Fukatsu T: **Complex endosymbiotic microbiota of the citrus psyllid *Diaphorina citri* (Homoptera: Psylloidea).** *Zool Science* 2000, **17:**983-989.
11. Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P: **Cospeciation of psyllids and their primary prokaryotic endosymbionts.** *App Environ Microbiol* 2000, **66:**2898-2905.
12. Grindle N, Tyner JJ, Clay K, Fuqua C: **Identification of *Arsenophonus*-type bacteria from the dog tick *Dermacentor variabilis*.** *J Invertebr Pathol* 2003, **83:**264-266.
13. Russell JA, Latorre A, Sabater-Munoz B, Moya A, Moran NA: **Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea.** *Mol Ecol* 2003, **12:**1061-1075.
14. Zchori-Fein E, Brown JK: **Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae).** *Ann Entomol Soc Am* 2002, **95:**711-718.
15. Thao MLL, Baumann P: **Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae).** *Curr Microbiol* 2004, **48:**140-144.
16. Dale C, Beeton M, Harbison C, Jones T, Pontes M: **Isolation, pure culture, and characterization of "Candidatus *Arsenophonus arthropodicus*, " an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*.** *App Environ Microbiol* 2006, **72:**2997-3004.
17. Dunn AK, Stabb EV: **Culture-independent characterization of the microbiota of the ant lion *Myrmeleon mobilis* (Neuroptera: Myrmeleontidae).** *App Environ Microbiol* 2005, **71:**8784-8794.
18. Allen JM, Reed DL, Perotti MA, Braig HR: **Evolutionary relationships of "Candidatus *Riesia* spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice.** *App Environ Microbiol* 2007, **73:**1659-1664.
19. Babendreier D, Joller D, Romeis J, Bigler F, Widmer F: **Bacterial community structures in honeybee intestines and their response to two insecticidal proteins.** *Fems Microbiol Ecol* 2007, **59:**600-610.
20. Trowbridge RE, Dittmar K, Whiting MF: **Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea).** *J Invertebr Pathol* 2006, **91:**64-68.
21. Hansen AK, Jeong G, Paine TD, Stouthamer R: **Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California.** *App Environ Microbiol* 2007, **73:**7531-7535.
22. Semetey O, Gatineau F, Bressan A, Boudon-Padieu E: **Characterization of a gamma-3 proteobacteria responsible for the syndrome "basses richesses" of sugar beet transmitted by *Pentastiridius* sp. (Hemiptera, Cixiidae).** *Phytopathology* 2007, **97:**72-78.
23. Šorfová P, Škeříková A, Hypša V: **An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of *Arsenophonus triatominarum*.** *Syst and App Microbiol* 2008, **31:**88-100.
24. Perotti MA, Allen JM, Reed DL, Braig HR: **Host-symbiont interactions of the primary endosymbiont of human head and body lice.** *Faseb Journal* 2007, **21:**1058-1066.
25. Sasaki-Fukatsu K, Koga R, Nikoh N, Yoshizawa K, Kasai S, Mihara M, Kobayashi M, Tomita T, Fukatsu T: **Symbiotic bacteria associated with stomach discs of human lice.** *App Environ Microbiol* 2006, **72:**7349-7352.
26. Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, Sasaki-Fukatsu K, Yoshizawa K, Dale C, Clayton DH: **Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies.** *Appl Environ Microbiol* 2007, **73:**6660-6668.
27. Herbeck JT, Degnan PH, Wernegreen JJ: **Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (gamma-proteobacteria).** *Mol Biol Evol* 2005, **22:**520-532.
28. Baumann P: **Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects.** *Annu Rev Microbiol* 2005, **59:**155-189.
29. Lefevre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A: **Endosymbiont phylogenesis in the Dryophthoridae weevils: Evidence for bacterial replacement.** *Mol Biol Evol* 2004, **21:**965-973.
30. Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P: **Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae* : A peculiar G+C content of an endocytobiotic DNA.** *J Mol Evol* 1998, **47:**52-61.
31. Galtier N, Gouy M: **Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis.** *Mol Biol Evol* 1998, **15:**871-879.
32. Tamura K: **Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases.** *Mol Biol Evol* 1992, **9:**678-687.
33. Bergsten J: **A review of long-branch attraction.** *Cladistics* 2005, **21:**163-193.
34. Hypša V, Křížek J: **Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura).** *Microb Ecology* 2007, **54:**242-251.
35. Dittmar K, Porter ML, Murray S, Whiting MF: **Molecular phylogenetic analysis of nycteribiid and streblid bat flies (Diptera: Brachycera, Calyptratae): Implications for host associations and phylogeographic origins.** *Mol Phyl Evol* 2006, **38:**155-170.
36. Sandstrom JP, Russell JA, White JP, Moran NA: **Independent origins and horizontal transfer of bacterial symbionts of aphids.** *Mol Ecol* 2001, **10:**217-228.
37. Takiya DM, Tran PL, Dietrich CH, Moran NA: **Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts.** *Mol Ecol* 2006, **15:**4175-4191.
38. Thao ML, Gullan PJ, Baumann P: **Secondary (gamma-Proteobacteria) endosymbionts infect the primary (beta-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts.** *App Environ Microbiol* 2002, **68:**3190-3197.
39. Werren JH: **Biology of *Wolbachia*.** *Annu Rev Entomol* 1997, **42:**587-609.
40. Heath BD, Butcher RDJ, Whitfield WGF, Hubbard SF: **Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism.** *Curr Biol* 1999, **9:**313-316.
41. Russell JA, Moran NA: **Horizontal transfer of bacterial symbionts: Heritability and fitness effects in a novel aphid host.** *App Environ Microbiol* 2005, **71:**7987-7994.
42. Mylvaganam S, Dennis PP: **Sequence heterogeneity between the 2 genes encoding 16S ribosomal-RNA from the halophilic archeabacterium *Haloarcula marismortui*.** *Genetics* 1992, **130:**399-410.
43. Wang Y, Zwang ZS, Ramanan N: **The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes.** *J Bacteriol* 1997, **179:**3270-3276.

44. Miller SR, Sunny A, Olson TL, Blankenship RE, Selker J, Wood M: **Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial cyanobacterial small-subunit rRNA gene.** *Proc Natl Acad Sci USA* 2005, **102:**850-855.
45. Wang Y, Zhang ZS: **Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes.** *Microbiology-Sgm* 2000, **146:**2845-2854.
46. Gogarten JP, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19:**2226-2238.
47. Lin CK, Hung CL, Chiang YC, Lin CM, Tsen HY: **The sequence heterogenicities among 16S rRNA genes of *Salmonella* serovars and the effects on the specificity of the primers designed.** *Int J Food Microbiol* 2004, **96:**205-214.
48. Marchandin H, Teyssier C, de Buochberg MS, Jean-Pierre H, Carriere C, Jumas-Bilak E: **Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: implications for phylogeny and taxonomy.** *Microbiology-Sgm* 2003, **149:**1493-1501.
49. Pettersson B, Bolske G, Thiaucourt F, Uhlen M, Johansson KE: **Molecular evolution of *Mycoplasma capricolum* subsp. capripneumoniae strains, based on polymorphisms in the 16S rRNA genes.** *J Bacteriol* 1998, **180:**2350-2358.
50. Yap WH, Zhang ZS, Wang Y: **Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon.** *J Bacteriol* 1999, **181:**5201-5209.
51. Stewart FJ, Cavanaugh CM: **Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria.** *J Mol Evol* 2007, **65:**44-67.
52. Thao ML, Baumann P: **Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts.** *App Environ Microbiol* 2004, **70:**3401-3406.
53. Dale C, Wang B, Moran N, Ochman H: **Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration.** *Mol Biol Evol* 2003, **20:**1188-1194.
54. Battistuzzi FU, Feijao A, Hedges SB: **A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land.** *Bmc Evol Biol* 2004, **4:**14.
55. Ochman H, Wilson AC: **Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes.** *J Mol Evol* 1987, **26:**74-86.
56. Rutschmann F: **Bayesian molecular dating using PAML/multidivtime. A step-by-step manual.** 2005 [http://www.plant.ch]. University of Zurich, Switzerland
57. Gaunt MW, Miles MA: **An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks.** *Mol Biol Evol* 2002, **19:**748-761.
58. Moran NA, Wernegreen JJ: **Lifestyle evolution in symbiotic bacteria: insights from genomics.** *Trends Ecol Evol* 2000, **15:**321-326.
59. Dale C, Plague GR, Wang B, Ochman H, Moran NA: **Type III secretion systems and the evolution of mutualistic endosymbiosis.** *Proc Natl Acad Sci USA* 2002, **99:**12397-12402.
60. Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ: **Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: Cospeciation of *Camponotus* species and their endosymbionts, Candidatus *Blochmannia*.** *Syst Biol* 2004, **53:**95-110.
61. Moran NA, Tran P, Gerardo NM: **Symbiosis and insect diversification: An ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes.** *App Environ Microbiol* 2005, **71:**8802-8810.
62. Clark MA, Moran NA, Baumann P, Wernegreen JJ: **Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence.** *Evolution* 2000, **54:**517-525.
63. Duron O, Gavotte L: **Absence of *Wolbachia* in nonfilariid worms parasitizing arthropods.** *Curr Microbiol* 2007, **55:**193-197.
64. Allen JM, Light JE, Perotti MA, Braig HR, Reed DL: **Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet.** *PLoS One* 2009, **4(3):**e4969.
65. Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstadter J, Hurst GD: **The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone.** *BMC Biol* 2008, **6(1):**27.
66. Baldo L, Werren JH: **Revisiting *Wolbachia* supergroup typing based on WSP: Spurious lineages and discordance with MLST.** *Curr Microbiol* 2007, **55:**81-87.
67. Casiraghi M, Bordenstein SR, Baldo L, Lo N, Beninati T, Wernegreen JJ, Werren JH, Bandi C: **Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree.** *Microbiology-Sgm* 2005, **151:**4015-4022.
68. Werren JH: ***Arsenophonus*.** In *Bergey's Manual of Systematic Bacteriology Volume 2*. Edited by: Garrity GM. New York: Springer-Verlag; 2004.
69. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nuc Acid Symp Series* 1999, **41:**95-98.
70. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17:**540-552.
71. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14:**817-818.
72. Goloboff PA, Farris JS, Nixon KC: **TNT.** *Cladistics-the International Journal of the Willi Hennig Society* 2004, **20:**84-84.
73. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52:**696-704.
74. Galtier N, Gouy M, Gautier C: **SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny.** *Comput Appl Biosci* 1996, **12:**543-548.
75. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W: **Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data.** *Genetics* 2002, **161:**1307-1320.

Paper 8:

Chrudimský, T., Husník, F., **Nováková, E.**, & Hypša, V. (2012). *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. *PLoS One*, *7*(7), e40354.

OPEN ACCESS Freely available online

PLoS one

Candidatus Sodalis melophagi sp. nov.: Phylogenetically Independent Comparative Model to the Tsetse Fly Symbiont *Sodalis glossinidius*

Tomáš Chrudimský[1]*, Filip Husník[1,2], Eva Nováková[1,2], Václav Hypša[1,2]

1 Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, **2** Institute of Parasitology, Biology Centre, Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic

Abstract

Bacteria of the genus *Sodalis* live in symbiosis with various groups of insects. The best known member of this group, a secondary symbiont of tsetse flies *Sodalis glossinidius*, has become one of the most important models in investigating establishment and evolution of insect-bacteria symbiosis. It represents a bacterium in the early/intermediate state of the transition towards symbiosis, which allows for exploring such interesting topics as: usage of secretory systems for entering the host cell, tempo of the genome modification, and metabolic interaction with a coexisting primary symbiont. In this study, we describe a new *Sodalis* species which could provide a useful comparative model to the tsetse symbiont. It lives in association with *Melophagus ovinus*, an insect related to tsetse flies, and resembles *S. glossinidius* in several important traits. Similar to *S. glossinidius*, it cohabits the host with another symbiotic bacterium, the bacteriome-harbored primary symbiont of the genus *Arsenophonus*. As a typical secondary symbiont, *Candidatus* Sodalis melophagi infects various host tissues, including bacteriome. We provide basic morphological and molecular characteristics of the symbiont and show that these traits also correspond to the early/intermediate state of the evolution towards symbiosis. Particularly, we demonstrate the ability of the bacterium to live in insect cell culture as well as in cell-free medium. We also provide basic characteristics of type three secretion system and using three reference sequences (16 S rDNA, *groEL* and *spaPQR* region) we show that the bacterium branched within the genus *Sodalis*, but originated independently of the two previously described symbionts of hippoboscoids. We propose the name *Candidatus* Sodalis melophagi for this new bacterium.

Citation: Chrudimský T, Husník F, Nováková E, Hypša V (2012) *Candidatus* Sodalis melophagi sp. nov.: Phylogenetically Independent Comparative Model to the Tsetse Fly Symbiont *Sodalis glossinidius*. PLoS ONE 7(7): e40354. doi:10.1371/journal.pone.0040354

Editor: Adam Driks, Loyola University Medical Center, United States of America

Received November 17, 2011; **Accepted** June 7, 2012; **Published** July 17, 2012

Funding: The study was supported by grants LC06073 and MSM6007665801 (Ministry of Education, Youth and Sports of the Czech Republic) and grant P505/10/1401 (Grant agency of the Czech Republic). Access to the MetaCentrum computing facilities was provided by "Projects of Large Infrastructure for Research, Development, and Innovations" LM2010005 (Ministry of Education, Youth, and Sports of the Czech Republic). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tomas.chrudimsky@gmail.com

Introduction

The genus *Sodalis* belongs to the symbiotic bacterial lineages that adopted several different types of symbiosis with their hosts, ranging from facultative commensals to obligate mutualists [1,2]. *Sodalis* and closely related bacteria were described from a broad spectrum of insect hosts including tsetse flies [3], weevils [4–7], chewing lice [8], hippoboscid louse flies [9], ants [10], scale insects [11], aphids [12], stinkbugs [13,14], and cerambycid beetles [15]. Also, some "secondary" symbionts of psyllids and mealybugs cluster with the *Sodalis* clade [16,17]. Within symbiotic Enterobacteriaceae, diversity of the *Sodalis* clade is comparable only with the genus *Arsenophonus* [18].

The first described, best known and most frequently investigated member of the genus is *S. glossinidius*, S-symbiont of tsetse flies [3]. Its significance for the host is still not clear, but a possible influence on the host longevity and resistance to trypanosomes has previously been suggested [19]. Several molecular analyses and genetic experiments made *S. glossinidius* an important model for investigating evolution and biology of symbiotic bacteria [20–22]. It has been employed in research of various biological traits, such as the structure and role of secretion systems [21], the function of the iron acquisition system [23] or the usage of the quorum sensing system [24]. Moreover, *S. glossinidius* proved to be among the few symbionts that could be maintained in *in vitro* culture in insect cells as well as in the cell-free media [3,25]. This feature has been attributed to the initial or intermediate state of the *S. glossinidius* shift towards symbiosis. One of the most frequently discussed topics in this respect is the state and function of the type three secretion system (TTSS) in this bacterium. Three different copies of the TTSS (SSR-1, SSR-2 and SSR-3) has been detected in *S. glossinidius* [26,27] and a possible role of the SSR-2 in invading host cells has been proposed [21].

Within the host, *S. glossinidius* constitutes part of a complex bacterial community which also contains P-symbiont *Wigglesworthia glossinidia* [28] and alphaproteobacterium *Wolbachia* [29,30]. Recent investigations show that the whole community may be even richer and contain an array of other bacteria [31]. Such complex host-symbiont systems provide a unique opportunity for comparing genomes in different states/modes of symbiosis and studying processes of their metabolic complementation [32,33]. For the *Sodalis-Wigglesworthia-Glossina* association,

complete genomes of both bacteria have been sequenced and annotated [26,34]. Even though each of the sequenced genomes came from different host species, an interaction between *Sodalis* and W*iggleswortbia* via thiamine synthesis could be detected by their comparison [27] and this view was further corroborated by an experimental approach [35] and sequencing of *Wigglesworthia* lineage from *Glossina morsitans* [36].

An establishment of a more complete picture of *Sodalis* genome evolution will require complete genome sequences for a more diverse array of *Sodalis* isolates, and of similar complex systems involving this bacterium. Although tsetse flies are the most important blood feeding brachycerans, there are several related groups of dipterans that display many similar features such as the feeding strategy, vivipary and transmission of trypanosomes.

A member of the genus *Sodalis*, phylogenetically independent on the tsetse symbiont, has already been described from hippoboscoid species *Craterina melbae* [9]. Here, we characterize a new member of the *Sodalis* lineage, inhabiting gut and other tissues of another hippoboscoid, *Melophagus ovinus*. The presence of several symbiotic bacteria in this species has long been known. According to morphological investigations of several researchers summarized by Paul Buchner [37], *Melophagus ovinus* contains symbiotic bacteria within enlarged epithelial cells of a specialized section of the midgut (bacteriome). This P-symbiont was recently characterized by molecular techniques as a member of *Arsenophonus* clade and is likely to play a role resembling that of *Wigglesworthia* in tsetse flies (Nováková et al., in prep). In addition, some of the authors recognized two other bacteria in the sheep keds. The first is *Bartonella melophagi* (originally described as *Rickettsia melophagi* and *Wolbachia melophagi*), which is localized extracellularly along microvilli of the midgut. The second type of bacterium described in Paul Buchner's work resembles *Candidatus* Sodalis melophagi as presented in this study: *"In the low zones of the midgut epithelium…, there are additional delicate bacteria, sometimes forming rather long filaments, which also must not be confused with the symbionts."* The whole system thus remarkably resembles the *Wigglesworthia-Sodalis* association in tsetse flies and can provide important data for a comparative study. In this study, we present a basic molecular and morphological characterization of the new *Sodalis* linage and overview the composition of its TTSS. We suggest the new name *Candidatus* Sodalis melophagi for this bacterium and extend the available *Sodalis* spectrum with three additional samples which allow for more precise phylogenetic characterization.

Results

Sequence Data

Sequences obtained by PCR for the investigated samples and their accession numbers are summarized in Table S1. The 16 S rDNA sequence from *Candidatus* Sodalis melophagi sp. nov. displayed 98.48% similarity to 16 S rDNA of *S. glossinidius*. Illumina assemblies produced preliminary draft sequences from which only selected gene regions were used here for the formal description and basic phylogenetic characterization. These regions, including all TTSS genes (Table S1) and *groEL* chaperonin, were of a high quality and did not contain any SNPs. In order to avoid assembly artifacts affecting the sequence accuracy of highly similar paralogous regions, a partial sequence for 16 S rRNA gene was obtained through Sanger sequencing as described above and was used for inferring phylogeny. Sanger sequenced *groEL* partial sequence was identical to the sequence acquired from Illumina data and we therefore used the full length *groEL* for phylogenetic reconstruction.

Type Three Secretion System (TTSS)

Candidatus Sodalis melophagi possesses only SSR-2 and SSR-3 copies of the TTSS in its genome, while SSR-1 is completely missing. Extensive BLASTX searches did not produce any significant hits for *S. glossinidius* SSR-1 either in Illumina contigs or raw reads. The gene order and content of SSR-2 and SSR-3 of *Candidatus* Sodalis melophagi is very similar, but not identical to that in *S. glossinidius* (Figure 1). The SSR-2 sequence comprises 11 protein coding genes: *orgAbA*, *prgKIH*, *spaSQPO*, *invF* and *hilA*; and 5 pseudogenes: *prgJ, sicA*, *spaR* and *invAG*. In comparison with *S. glossinidius*, it lacks 4 genes: *invB, invC* and *spaLMN*. The SSR-3 comprises 29 protein coding genes: *ssrAB*, *ssaBCDEGHIJKLMN OPQRSTUV*, *sseABCDE*, *sscB*, a protein similar to locus SG1296 of *S. glossinidius*, and a single pseudogene: *sscA*. Sequences of *Candidatus* Sodalis melophagi TTSS genes were deposited in GenBank as a part of two annotated contigs for each of the islands (Table S1).

Phylogenetic Analyses

The lengths of individual matrices and numbers of variable positions are summarized in Table S2. All phylogenetic trees clearly indicate that the novel bacterium belongs to the genus *Sodalis* (Figures 2, 3, 4, Figure S1, Figure S2). In the trees derived from 16 S rRNA gene sequence data and amino acid sequence of *groEL*, *Candidatus* Sodalis melophagi clusters within a large polytomy and its precise position within the genus is thus uncertain (Figures 2–3A, Figure S1, Figure S2). However, even this unresolved topology excludes its relationship to any of the two other hippoboscoid-derived *Sodalis* members. This conclusion is further supported by the nucleotide matrix for *groEL* (Figure 3B) with a reduced sampling and the *spaPQR* concatenate (Figure 4). Although the exact topology slightly varies with the methods and parameters of the analysis, the three hippoboscoid lineages always form a polyphyletic/paraphyletic assemblage.

In vitro Culture

Bacterial colonies were clearly visible after 8 days of cultivation. Colonies were white, raised, and circular with entire edges. Their size was irregular ranging from 0.5 to 1 mm. The variable size of the colonies was almost certainly not due to new mutations, since subculturing on fresh plates yielded the same variability for each individual colony. The irregular colony morphology was described also in case of a type strain $M1^T$ of the *Sodalis glossinidius* [3] and is probably population dependent as is the case in other microaerophilic bacteria [38].

The type strain CZ^T was established by isolation of a single bacterial colony and was used for C6/36 cells infection. In the C6/36 cell culture, bacteria were predominantly attached to cell surface or free in the medium, but they were also observed inside the cells. Genomic DNA purified from bacterial colonies of *Candidatus* Sodalis melophagi CZ^T was used as a template for Illumina mate pair sequencing.

Microscopy

Under electron microscopy, *Sodalis* cells within the host tissue corresponded well to their light microscope characteristics (Figure 5A). They appeared as rods reaching from approx. 1 to 4 μm, depending on the angle of the section, and were located mainly at the periphery of the bacteriome, sometimes in close association with the P-symbionts (Figures 5B,C).

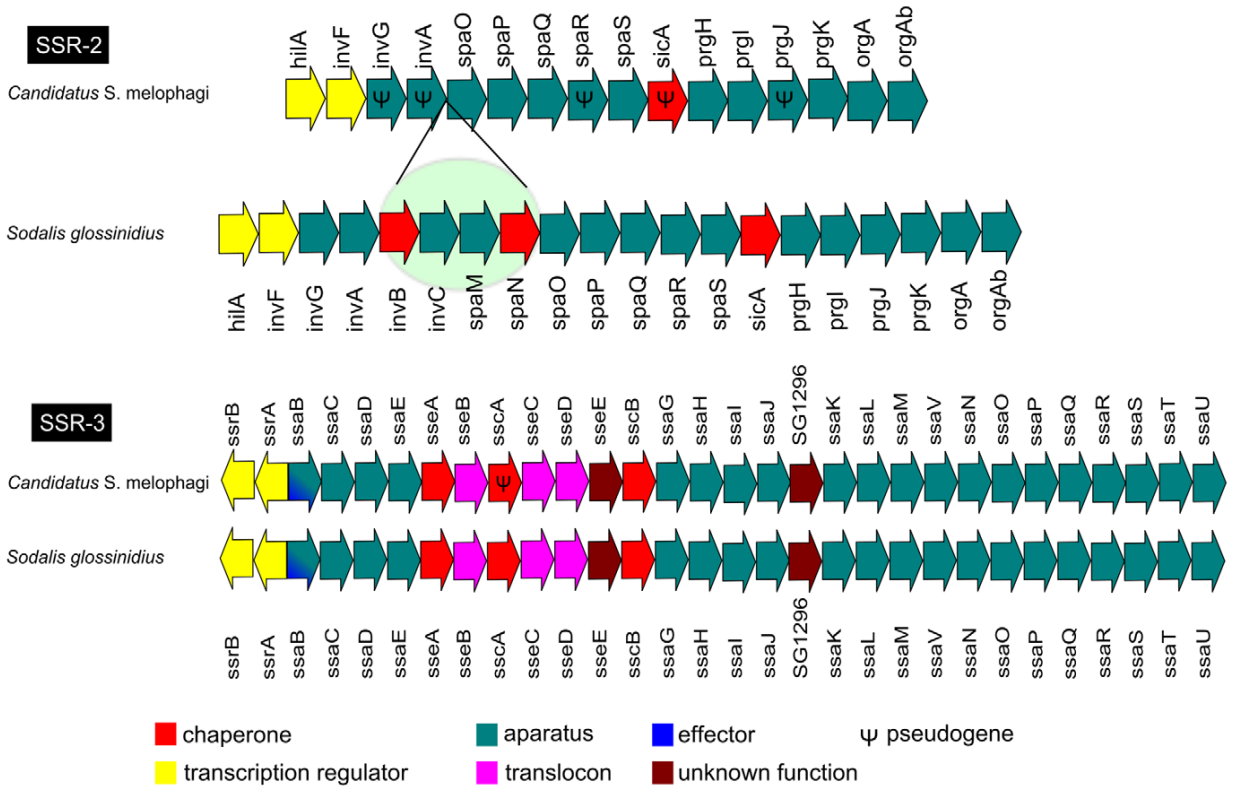


Figure 1. Composition of SSR-2 and SSR-3 copies of *Candidatus* Sodalis melophagi TTSS. Ψ – putative pseudogene (see Material and Methods).
doi:10.1371/journal.pone.0040354.g001

Description of *Candidatus* Sodalis Melophagi sp. nov

Candidatus Sodalis melophagi (me.lo.phag.i; *melophagi* of the genus *Melophagus*).

Candidatus Sodalis melophagi is a Gram-negative bacterium isolated from the sheep ked *Melophagus ovinus*. The bacteria were detected in heamolymph and bacteriome. The cells are rod-shaped non-motile and non-spore forming under laboratory conditions. The average length is 1.9 μm, however, bacteria ranging from 0.6 to 5 μm in length were observed. The bacteria grow intracelullarly and extracellularly in presence of *Aedes albopictus* cell line C6/36 under aerobic conditions. They can also be cultivated axenically both in liquid and solid media containing enzymatically digested proteins as a source of nitrogen, with the addition of blood under the microaerophilic atmosphere (5% O_2 balanced with carbon dioxide) at 27°C.

Discussion

Remarkably, the main biological traits of *Candidatus* Sodalis melophagi and its relation to the host resemble those of *Sodalis glossinidius* in tsetse flies. In both cases the hosts are exclusively blood feeding organisms reproducing by vivipary and their *Sodalis* plays a role of an "accessory" symbiont accompanying phylogenetically distant P-symbiont (*Wigglesworthia* in tsetse flies and *Arsenophonus* in *Melophagus*). This status of *Candidatus* Sodalis melophagi is strongly indicated by phylogenetic characteristics as well as location and morphology revealed by electron microscopy. Similar to *S. glossinidius*, the rod-shaped cells of *Candidatus* Sodalis melophagi are not restricted to the bacteriocytes but they infect various non-specialized cells and can be cultivated from haemolyph. It is therefore interesting to note that according to phylogenetic analyses its symbiosis with the host originated independently of the *S. glossinidius* as well as of the third hipoboscoid lineage, the symbiont of *Craterina melbae*. Although exact position of *Candidatus* Sodalis melophagi within the *Sodalis* clade varies with method and taxon sampling, it never clusters with any of the two other hippoboscoid-derived lineages. Within the *Sodalis* clade, such host-symbiont incongruence is not unique to the hippoboscoid-derived lineages. It is also expressed by the distribution of several other samples, particularly those from coleopteran and homopteran hosts (Figure 2, Figure S1) and clearly demonstrates the capability of *Sodalis* to spread by horizontal transfers. In contrast, the close relationships of the two new 16 S rDNA sequences obtained from *Ornithomya* with the *Sodalis* previously described from *Craterina melbae* suggests that these two hippoboscoid genera may share the same symbiotic lineage inherited from their common ancestor. However, the sampling of *Sodalis* lineages available from other insect groups is highly incomplete. In addition, the position of some sequences, such as *Sodalis* from *Sitophilus rugicollis* is affected by long-branch attraction due to the high AT content; instability of this lineage has been found in previous analyses [9]. Thus, although we extended the sampling with additional three sequences, any interpretations of the modes and mechanism of *Sodalis* transmission can under the current circumstances be only working hypotheses rather than serious conclusions based on the data.

While the draft genomic data of *Candidatus* Sodalis melophagi is currently under investigation, we preliminarily analyzed the

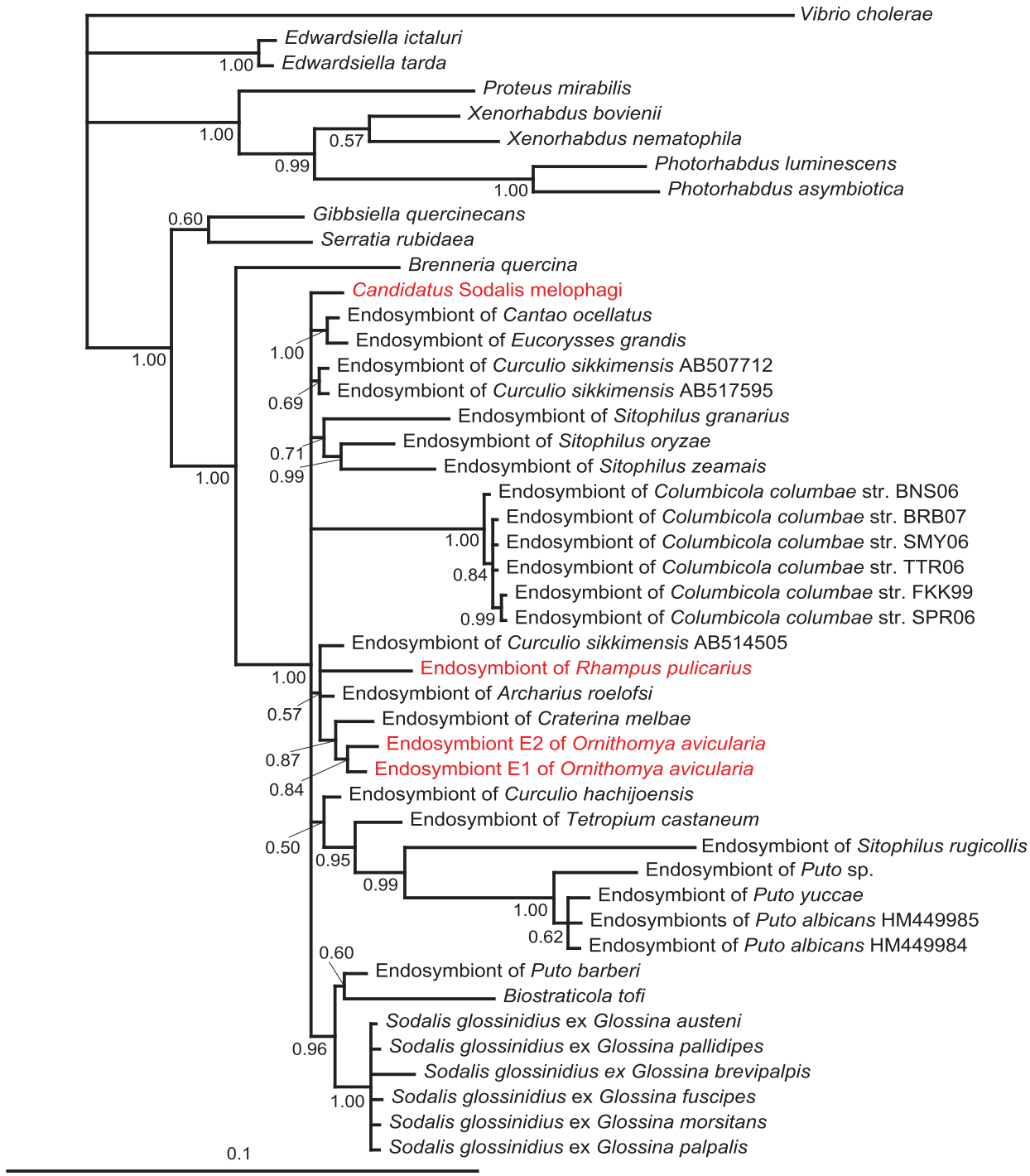


Figure 2. 16 S rDNA tree derived by BI analysis in MrBayes. Posterior probabilities are indicated by the numbers at the nodes. New *Sodalis* lineages added in this study are printed in red.
doi:10.1371/journal.pone.0040354.g002

composition and structure of the TTSS. Significance of this system in symbiosis evolution has previously been suggested [39] and investigated in several symbiotic lineages [21,40–43]. These genes are also among the few sequences that are currently available and can be compared between different *Sodalis* taxa. In *S. glossinidius*, experimental work indicated that TTSS of SPI1 type from *Salmonella* (later designated as SSR-2) is essential for entering the host cell. Presence and apparent functionality of this system was subsequently confirmed in another *Sodalis*, the primary endosymbiont of *Sitophilus zeamais* (SZPE) [42]. However, further studies

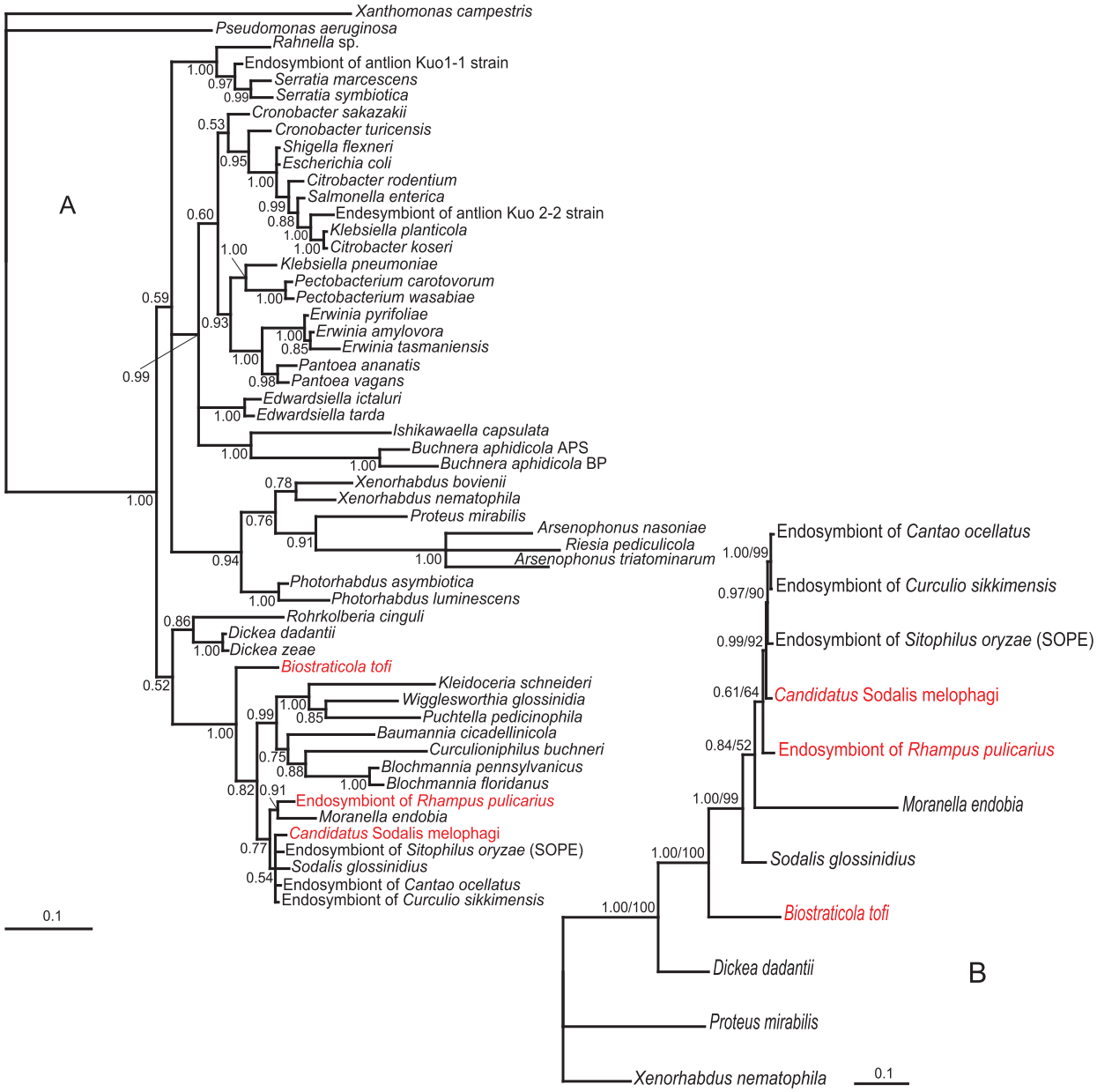


Figure 3. Phylogenetic trees derived from *groEL* matrices by BI in MrBayes. New *Sodalis* lineages added in this study are printed in red. **A**: The tree inferred from aminoacid matrix. Posterior probabilities are indicated by the numbers at the nodes. **B**: The tree inferred from nucleotide matrix restricted taxonomically to the *Sodalis* branch. The numbers at the nodes show the posterior probabilities and bootstrap values from the identical topology obtained by ML in PhyML.
doi:10.1371/journal.pone.0040354.g003

revealed another two copies of TTSS in *S. glossinidius* genome, SSR-1 closely related to the Ysa system of *Yersinia*, and SSR-3 related to *Salmonella* SPI2 [26,27,40]. The secretion systems in *Candidatus* Sodalis melophagi show clear similarity to *S. glossinidius* genes but the whole machinery is much less complex; SSR-1 is completely missing and SSR-2 is highly eroded. However, the form corresponding to the SSR-3 is complete and possibly functional. Its structure and gene content is highly similar to that in *S. glossinidius*, except for a pseudogenized state of *sscA*, encoding a putative chaperone of secreted protein SseC [44].

The differences between TTSS in *S. glossinidius* and *Candidatus* Sodalis melophagi pose an interesting question about the origin, role and significance of TTSS and its different copies in the genus *Sodalis*. As the two compared lineages, *Candidatus* Sodalis melophagi and *S. glossinidius* are not closely related and the whole *Sodalis* tree is currently undersampled, it is impossible to hypothesize whether SSR-1 was lost in the former one or acquired by the latter one after diversification of their ancestor. The significance of TTSS in the two lineages is even more difficult to assess. In *Candidatus* Sodalis melophagi, the SSR-3 alone or

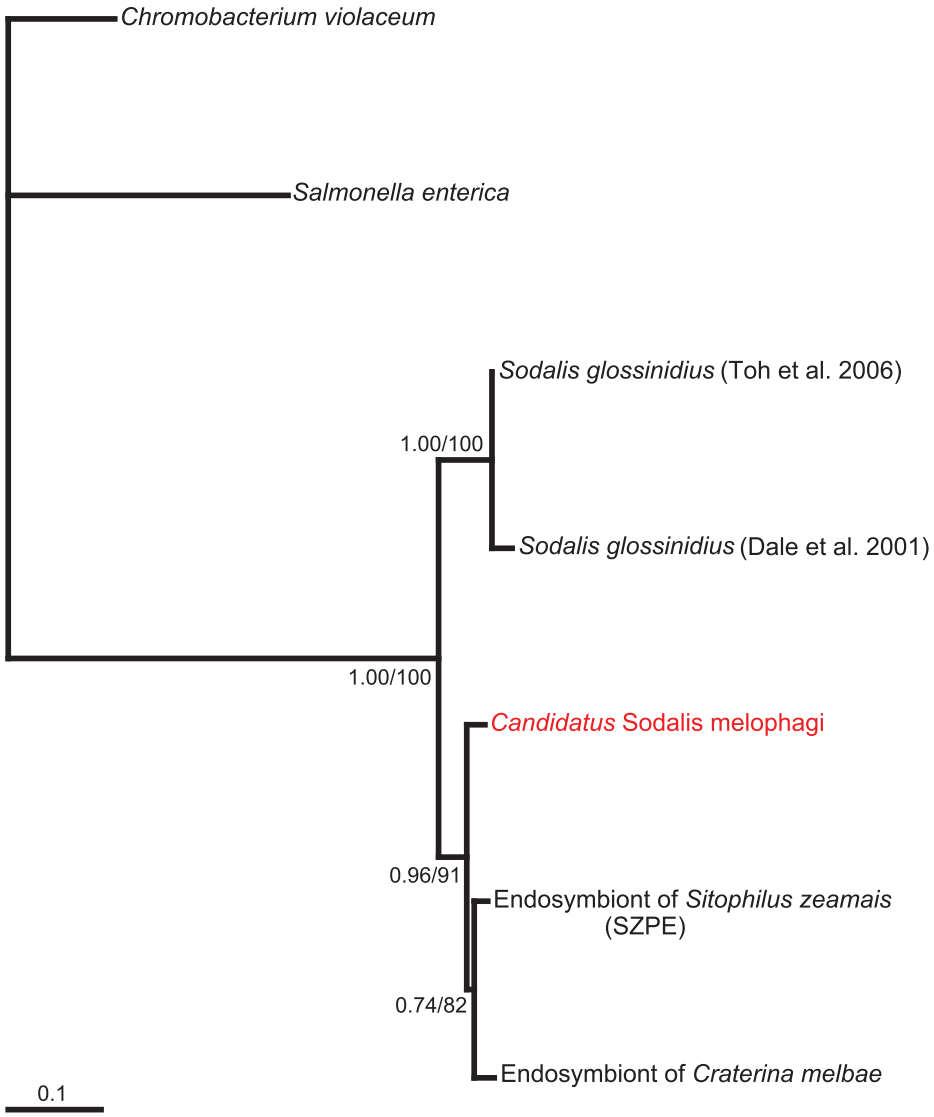


Figure 4. Phylogenetic tree derived from *spaPQR* region by BI in MrBayes. The numbers at the nodes show the posterior probabilites and bootstrap values from the identical topology obtained by ML in PhyML. New *Sodalis* lineages added in this study are printed in red.
doi:10.1371/journal.pone.0040354.g004

possibly in synergy with the retained functional genes of SSR-2, seems to be sufficient for maintenance of the intracellular lifestyle. This view, however, is based solely on the observed presence/absence of the genes and their comparison to the experimental results from *S. glossinidius* [21]. It should be taken into account that other related systems, such as flagellar export apparatus, can participate in the host cell invasion. Despite the knockout experiments, the situation is similarly unclear in *S. glossinidius* where the presence of SSR-3 has not been known at the time of the experimental work and its significance could not be investigated. From the genomic point of view, *S. glossinidius* has so far been the only member of the genus for which a detailed characterization of the genome and some metabolic capabilities is available. The quality of *Candidatus* Sodalis melophagi paired-end data retrieved from the bacteriome sequencing and additional mate pair data recently obtained from the pure culture suggest that a draft genome of this symbiont could be established and used for further comparisons. Such analysis, comparing reduction, structure, and possible adaptive changes of independent but closely related bacteria from two hosts with similar but unique biology will provide important insight into the symbiogenetic processes. In respect to the future work, it may help to discriminate between the random and symbiosis-associated modifications and indicate candidate genes for a more detailed investigation.

Materials and Methods

Ethic Statement

All field studies did not involve protected or endangered organisms. They were not performed on privately-owned or protected locations and were performed according to the law of the Czech Republic.

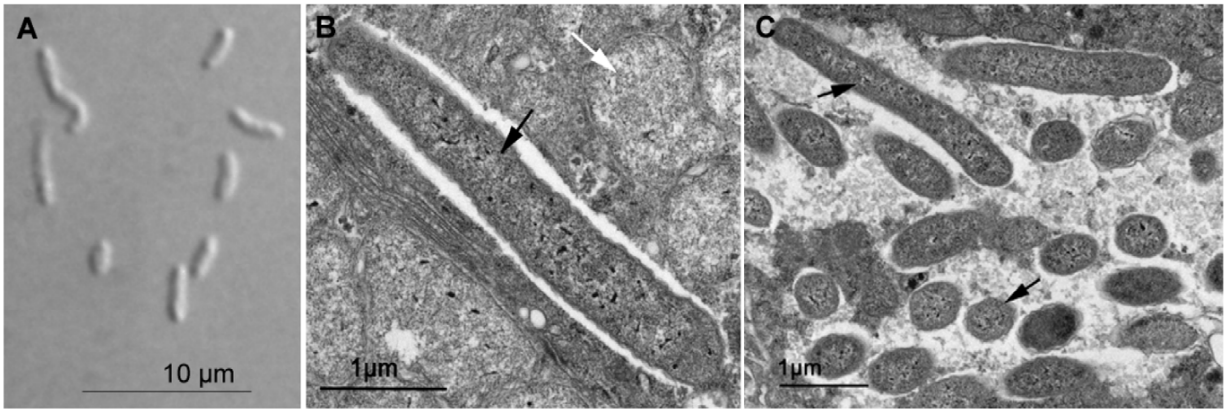


Figure 5. Morphology and ultrastructure of *Candidatus* Sodalis melophagi. A: *In vitro* cell culture in Nomarski contrast. **B**, **C**: Cells of *Candidatus* Sodalis melophagi in bacteriome. Black arrows – cells of *Candidatus* Sodalis melophagi, white arrows – cells of the primary endosymbiont of the genus *Arsenophonus*.
doi:10.1371/journal.pone.0040354.g005

Insect Samples, DNA Extraction, Sequencing and Data Assembly

Adults of *Melophagus ovinus* were obtained commercially from a licensed family sheep farm at Krásetín, Czech Republic. A midgut region with the bacteriome was dissected into the phosphate buffered saline (PBS) and total genomic DNA was extracted from each single adult by QIAamp DNA Micro Kit (Qiagen). DNA concentration was determined using NanoDrop2000 (Thermo Scientific) and its quality was assessed in 2% agarose gel using standard electrophoresis. PCR was carried out as described previously [18] with bacterial primers for 16 S rDNA and *groEL* (Table S3). The same procedure was followed for three additional samples. Two of them were *Sodalis* bacteria from other insect hosts, *Ornithomya avicularia* (Diptera, Hippoboscidae) and *Rhampus pulicarius* (Coleoptera, Curculionidae), and the third was *Biostraticola tofi* strain BF36^T, a free-living bacterium supposedly closely related to *Sodalis* [45]. Laboratory culture of *B. tofi* was obtained from the DSMZ microorganism collection (Germany), *R. pulicarius* was collected near České Budějovice (Czech Republic); according to the law of the Czech Republic, no permits are required for collection of this organism. Samples of *O. avicularia* were provided by Department of Zoology (University of South Bohemia); the collections were done during the ornithological research performed in accordance with the law of the Czech Republic.

We used 3 μg of genomic DNA isolated from *M. ovinus* bacteriome as a template for an Illumina paired-end library with an insert size of 300 bp. Library construction and sequencing on one lane in a 100 bp run was carried out at Keck Microarray Resource, West Haven CT, USA. Reads obtained underwent adaptor and quality trimming, and were further processed using two different approaches. First, the reads with significant BLASTN [46] hit to *Sodalis glossinidius* genome sequence (NC_007712) were filtered and the retrieved subset was assembled using CLC Genomic Workbench (CLC bio A/S) with parameters set to the following values: similarity 0.9, length fraction 0.9, costs for deletion/insertion/mismatch 3. Second, de-novo assembly of all the processed reads into contigs was done on CLC Genomic Workbench (CLC bio A/S) under the same parameter setting. Contigs from this assembly were binned based on their average coverage and BlastX [46] hits against available bacterial genomes. Since some of the sequences were used in this study for phylogenetic reconstruction, we checked the accuracy of the Illumina-derived sequences by an independent Sanger sequencing of the *groEL* gene.

Type Three Secretion System (TTSS) Annotation

Contigs spanning corresponding genes to TTSS islands from *Sodalis glossinidius* were retrieved based on the Blast X [46] results from both assemblies. In order to obtain a single contig for each of the TTSS islands, gaps were closed using targeted Sanger sequencing (Table S3). Mapped sequences were checked for presence of single nucleotide polymorphisms (SNPs). ORF prediction was done using CLC Genomic Workbench (CLC bio A/S) with the minimum length set to 30 AA. Gene annotation based on similarity confirmed by BLAST searches was performed manually in the same software. All genes that contained frame shift mutation or stop mutations were tentatively classified as pseudogenes. The *hilA* previously classified as pseudogene [26,27] was annotated as functional based on recent experimental work [47].

Phylogenetic Analyses

We used three different regions for the phylogenetic reconstruction, 16 S rDNA, *groEL* and the TTSS region consisting of the *spaP-spaQ-spaR* genes (Tables S4, S5, S6). The *groEL* amino acid dataset was aligned using ClustalW algorithm in BioEdit with default parameters [48] and all ambiguously aligned sites were removed from the further analyses. To gain more precise phylogenetic resolution within the *Sodalis* branch, we narrowed the taxon sampling and prepared an additional matrix using nucleotide sequences for the *groEL* gene. The matrix was aligned as described above and edited manually. The 16 S rDNA and the *spaPQR* dataset were aligned in the Mafft program [49], using the E-INS-i strategy with default parameters and manually edited in BioEdit [48]. For 16 S rDNA, the ambiguously aligned positions were eliminated in Gblocks [50].

Two approaches were used to infer phylogenetic trees, maximum likelihood (ML) and Bayesian inference (BI). For ML we used PhyML v3.0 [51,52] with the SPR search algorithm. BI was performed in MrBayes 3.1.2 [53,54] with five million generations and tree sampling every 100 generations. AWTY [55] was used to check the MCMC convergence and determine burn-in. Evolutionary substitution models for proteins and DNA were selected by ProtTest 2.4 [56] and for DNA by jModelTest 0.1.1 [57], respectively. For DNA sequences, the General Time

Reversible (GTR) model was used with an estimated proportion of invariable sites (I) and heterogeneity of evolutionary rates modeled by the eight substitution rate categories of the gamma (Γ) distribution and the gamma shape parameter (alpha) estimated from the data. LG+I+Γ was determined as the best fitting model for the amino acid *groEL* dataset, and it was used for the ML analyses. Since this model is not implemented in MrBayes, we replaced it with WAG+I+Γ for the BI analysis. JTT+I+Γ model was used for the *spaPQR* concatenate in both BI and ML analyses.

Cultivation of *Candidatus* Sodalis Melophagi sp. nov

The insects were surface sterilized using 96% ethanol and heamolymph was collected into a 1.5 ml eppendorf tube containing 500 µl of Mitsuhashi-Maramorosch (MM) medium [58] with 20% heat inactivated foetal bovine serum (FBS). The tube was incubated overnight at 27°C without shaking, a slightly modified protocol of Matthew et al. [25] was followed. Liquid culture was plated onto 10% sheep blood MMI plates solidified by 1% agar. The medium was supplemented with 100 µg/ml Polymyxin B and 10 mg/ml Amphotericin B to prevent contamination by Gram-negative non-symbiotic bacteria [59] and fungi. The plates were incubated at 27°C in a microaerobic atmosphere generated by the Campygen pack system (Oxoid) producing 5% of oxygen balanced with carbon dioxide. Bacteria were further inoculated into flasks containing C6/36 mosquito cells [60] (LGC Standards, Czech Republic) in MM medium with 20% heat inactivated FBS.

Microscopy

A three day old liquid culture of C6/36 cells infected with *Candidatus* Sodalis melophagi was used for microscopic examination. Cells were harvested and fixed in 4% formaldehyde in PBS and observed under the BX53 microscope (Olympus) using Nomarski contrast. For electron microscopy, the midgut region containing bacteriome was dissected directly into a 2.5% glutaraldehyde in 0.1 M phosphate buffer and prefixed at 4°C overnight. The tissue was then postfixed at 4°C for 60 min with 2% osmium tetroxide in phosphate buffer. After dehydration through ethanol series, the samples were embedded in Spurr resin. Ultrathin sections were stained with uranyl acetate and lead citrate and examined in transmission electron microscope JEOL JEM-1010.

Supporting Information

Figure S1 16 S rDNA tree derived by ML method in PhyML. Bootstrap values are indicated by the numbers at the nodes. New *Sodalis* lineages added in this study are printed in red. Values under 50 are not shown.
(EPS)

Figure S2 Phylogenetic trees derived from *groEL* aminoacid matrix under ML in PhyML. Bootstrap values are indicated by the numbers at the nodes. New *Sodalis* lineages added in this study are printed in red. Values under 50 are not shown.
(EPS)

Table S1 List of sequences acquired in this study.
(DOC)

Table S2 Characteristics of particular datasets.
(DOC)

Table S3 List of primers used in this study.
(DOC)

Table S4 List of 16 S rDNA sequences used for phylogenetic inference.
(DOC)

Table S5 List of *groEL* sequences used for phylogenetic inference.
(DOC)

Table S6 List of *spaPQR* sequences used for phylogenetic inference.
(DOC)

Acknowledgments

We are grateful to Nancy A. Moran and Keck Microarray Resource for technical assistance with the genome sequencing.

Author Contributions

Wrote the paper: TC FH EN VH. Designed the study: EN VH. Established in vitro culture: TC FH. Generated and assembled the genomic data: EN. Performed the light and electron microscopy characterization: TC FH VH. Performed phylogenetic analyses: TC FH. Annotated and analyzed the TTSS sequences: TC.

References

1. Snyder AK, McMillen CM, Wallenhorst P, Rio RV (2011) The phylogeny of *Sodalis*-like symbionts as reconstructed using surface-encoding loci. FEMS Microbiol Lett 317: 143–151.
2. Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P (1998) Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G + C content of an endocytobiotic DNA. J Mol Evol 47: 52–61.
3. 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 Pt 1: 267–275.
4. Toju H, Hosokawa T, Koga R, Nikoh N, Meng XY, et al. (2010) "*Candidatus* Curculioniphilus buchneri," a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. Appl Environ Microbiol 76: 275–282.
5. Toju H, Fukatsu T (2011) Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. Mol Ecol 20: 853–868.
6. Conord C, Despres L, Vallier A, Balmand S, Miquel C, et al. (2008) Long-term evolutionary stability of bacterial endosymbiosis in curculionoidea: additional evidence of symbiont replacement in the dryophthoridae family. Mol Biol Evol 25: 859–868.
7. Campbell BC, Bragg TS, Turner CE (1992) Phylogeny of symbiotic bacteria of four weevilspecies (coleoptera: curculionidae) based on analysis of 16 S ribosomal DNA. Insect Biochem Mol Biol 22: 415–421.
8. Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, et al. (2007) Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies. Appl Environ Microbiol 73: 6660–6668.
9. Nováková E, Hypša V (2007) A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiol Lett 269: 131–135.
10. Sameshima S, Hasegawa E, Kitade O, Minaka N, Matsumoto T (1999) Phylogenetic comparison of endosymbionts with their host ants based on molecular evidence. Zoolog Sci 16: 993–1000.
11. Gruwell ME, Hardy NB, Gullan PJ, Dittmar K (2010) Evolutionary relationships among primary endosymbionts of the mealybug subfamily phenacoccinae (hemiptera: Coccoidea: Pseudococcidae). Appl Environ Microbiol 76: 7521–7525.
12. Burke GR, Normark BB, Favret C, Moran NA (2009) Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. Appl Environ Microbiol 75: 5328–5335.
13. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, et al. (2010) Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. Appl Environ Microbiol 76: 3486–3494.
14. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, et al. (2011) Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera: Scutelleridae). Zoolog Sci 28: 169–174.
15. Grünwald S, Pilhofer M, Höll W (2010) Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles (Coleoptera: Cerambycidae). Syst Appl Microbiol 33: 25–34.

16. Thao ML, Clark MA, Baumann L, Brennan EB, Moran NA, et al. (2000) Secondary endosymbionts of psyllids have been acquired multiple times. Curr Microbiol 41: 300–304.
17. Thao ML, Gullan PJ, Baumann P (2002) Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl Environ Microbiol 68: 3190–3197.
18. Nováková E, Hypša V, Moran NA (2009) *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol 9: 143.
19. Dale C, Welburn SC (2001) The endosymbionts of tsetse flies: manipulating host-parasite interactions. Int J Parasitol 31: 628–631.
20. Pontes MH, Dale C (2011) Lambda red-mediated genetic modification of the insect endosymbiont *Sodalis glossinidius*. Appl Environ Microbiol 77: 1918–1920.
21. Dale C, Young SA, Haydon DT, Welburn SC (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci U S A 98: 1883–1888.
22. 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.
23. Runyen-Janecky LJ, Brown AN, Ott B, Tujuba HG, Rio RV (2010) Regulation of high-affinity iron acquisition homologues in the tsetse fly symbiont *Sodalis glossinidius*. J Bacteriol 192: 3780–3787.
24. Pontes MH, Babst M, Lochhead R, Oakeson K, Smith K, et al. (2008) Quorum sensing primes the oxidative stress response in the insect endosymbiont, *Sodalis glossinidius*. PLoS One 3: e3541.
25. Matthew CZ, Darby AC, Young SA, Hume LH, Welburn SC (2005) The rapid isolation and growth dynamics of the tsetse symbiont *Sodalis glossinidius*. FEMS Microbiol Lett 248: 69–74.
26. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, et al. (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res 16: 149–156.
27. Belda E, Moya A, Bentley S, Silva FJ (2010) Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics 11: 449.
28. 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.
29. Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, et al. (2000) Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. Med Vet Entomol 14: 44–55.
30. O'Neill SL, Gooding RH, Aksoy S (1993) Phylogenetically distant symbiotic microorganisms reside in *Glossina* midgut and ovary tissues. Med Vet Entomol 7: 377–383.
31. 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.
32. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol 4: e188.
33. McCutcheon JP, Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci U S A 104: 19392–19397.
34. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, et al. (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet 32: 402–407.
35. Snyder AK, Deberry JW, Runyen-Janecky L, Rio RV (2010) Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc Biol Sci 277: 2389–2397.
36. Rio RV, Symula RE, Wang J, Lohs C, Wu YN, et al. (2012) Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: glossinidae) obligate symbiont *Wigglesworthia*. MBio 3.
37. Buchner P (1965) Endosymbiosis of animals with plant microorganisms. New York: Interscience Publishers.
38. Krieg NR, Hoffman PS (1986) Microaerophily and oxygen toxicity. Annu Rev Microbiol 40: 107–130.
39. Coombes BK (2009) Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. Trends Microbiol 17: 89–94.
40. Dale C, Jones T, Pontes M (2005) Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. Mol Biol Evol 22: 758–766.
41. Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev 62: 379–433.
42. Dale C, Plague GR, Wang B, Ochman H, Moran NA (2002) Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc Natl Acad Sci U S A 99: 12397–12402.
43. Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H (2005) The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. Proc Natl Acad Sci U S A 102: 16919–16926.
44. Cirillo DM, Valdivia RH, Monack DM, Falkow S (1998) Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. Mol Microbiol 30: 175–188.
45. Verbarg S, Fruhling A, Cousin S, Brambilla E, Gronow S, et al. (2008) *Biostraticola tofi* gen. nov., spec. nov., a novel member of the family Enterobacteriaceae. Curr Microbiol 56: 603–608.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
47. Pontes MH, Smith KL, De Vooght L, Van Den Abbeele J, Dale C (2011) Attenuation of the sensing capabilities of PhoQ in transition to obligate insect-bacterial association. PLoS Genet 7: e1002349.
48. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser: 95–98.
49. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33: 511–518.
50. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552.
51. Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol 537: 113–137.
52. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.
53. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.
54. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
55. Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24: 581–583.
56. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105.
57. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
58. Mitsuhashi JM, Maramorosch K (1964) Leafhopper tissue culture: embryonic, nymphal and imaginal tissue from aseptic insects. Contribution from Boyce Thompson Institute: 435–460.
59. Dale C, Beeton M, Harbison C, Jones T, Pontes M (2006) Isolation, pure culture, and characterization of "*Candidatus* Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. Appl Environ Microbiol 72: 2997–3004.
60. Igarashi A (1978) Isolation of a Singh's *Aedes albopictus* cell clone sensitive to Dengue and Chikungunya viruses. J Gen Virol 40: 531–544.

Paper 9:

Šochová, E., Husník, F., **Nováková, E.**, Halajian, A., & Hypša, V. (2017). *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ*, *5*, e4099.

Arsenophonus and *Sodalis* replacements shape evolution of symbiosis in louse flies

Eva Šochová[1], Filip Husník[2,3], Eva Nováková[1,3], Ali Halajian[4] and Václav Hypša[1,3]

[1] Department of Parasitology, University of South Bohemia, České Budějovice, Czech Republic
[2] Department of Molecular Biology, University of South Bohemia, České Budějovice, Czech Republic
[3] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic
[4] Department of Biodiversity, University of Limpopo, Sovenga, South Africa

ABSTRACT

Symbiotic interactions between insects and bacteria are ubiquitous and form a continuum from loose facultative symbiosis to greatly intimate and stable obligate symbiosis. In blood-sucking insects living exclusively on vertebrate blood, obligate endosymbionts are essential for hosts and hypothesized to supplement B-vitamins and cofactors missing from their blood diet. The role and distribution of facultative endosymbionts and their evolutionary significance as seeds of obligate symbioses are much less understood. Here, using phylogenetic approaches, we focus on the Hippoboscidae phylogeny as well as the stability and dynamics of obligate symbioses within this bloodsucking group. In particular, we demonstrate a new potentially obligate lineage of *Sodalis* co-evolving with the Olfersini subclade of Hippoboscidae. We also show several likely facultative *Sodalis* lineages closely related to *Sodalis praecaptivus* (HS strain) and suggest repeated acquisition of novel symbionts from the environment. Similar to *Sodalis*, *Arsenophonus* endosymbionts also form both obligate endosymbiotic lineages co-evolving with their hosts (Ornithomyini and Ornithoica groups) as well as possibly facultative infections incongruent with the Hippoboscidae phylogeny. Finally, we reveal substantial diversity of *Wolbachia* strains detected in Hippoboscidae samples falling into three supergroups: A, B, and the most common F. Altogether, our results prove the associations between Hippoboscoidea and their symbiotic bacteria to undergo surprisingly dynamic, yet selective, evolutionary processes strongly shaped by repeated endosymbiont replacements. Interestingly, obligate symbionts only originate from two endosymbiont genera, *Arsenophonus* and *Sodalis*, suggesting that the host is either highly selective about its future obligate symbionts or that these two lineages are the most competitive when establishing symbioses in louse flies.

Subjects Biodiversity, Bioinformatics, Entomology, Evolutionary Studies
Keywords *Arsenophonus*, *Sodalis*, *Wolbachia*, Louse flies, Replacements, Phylogeny

Submitted 1 August 2017
Accepted 6 November 2017
Published 11 December 2017

Corresponding author
Eva Šochová, sochova.e@seznam.cz

Academic editor
Mikhail Gelfand

Additional Information and Declarations can be found on page 15

DOI **10.7717/peerj.4099**

BACKGROUND

Symbiotic associations are widespread among animals and bacteria and often considered to undergo a common evolution as a holobiont (*Zilber-Rosenberg & Rosenberg, 2008*). The host and symbiont are either fully dependent on each other for reproduction and survival

How to cite this article Šochová et al. (2017), *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. **PeerJ 5:e4099; DOI 10.7717/peerj.4099**

(obligate symbiosis) or not (facultative symbiosis), but in reality there is a gradient of such interactions (*Moran, McCutcheon & Nakabachi, 2008*). Any establishment of a symbiotic association brings not only advantages, but also several challenges to both partners. Perhaps the most crucial is that after entering the host, the endosymbiont genome tends to decay due to population genetic processes affecting asexual organisms with small effective population sizes (*Moran, 1996*) and the host is becoming dependent on such a degenerating symbiont (*Koga et al., 2007*; *Pais et al., 2008*). Since symbionts are essential for the host, the host can try to escape from this evolutionary 'rabbit hole' by an acquisition of novel symbionts or via endosymbiont replacement and supplementation (*Bennett & Moran, 2015*). This phenomenon, known in almost all insect symbiotic groups, was especially studied in the sap-feeding group Hemiptera (*Sudakaran, Kost & Kaltenpoth, 2017*), while only few studies were performed from blood-sucking groups.

Blood-sucking insects, living exclusively on vertebrate blood, such as sucking lice (*Allen et al., 2007*; *Hypša & Křížek, 2007*; *Fukatsu et al., 2009*; *Allen et al., 2016*), bed bugs (*Hypša & Aksoy, 1997*; *Hosokawa et al., 2010*; *Nikoh et al., 2014*), kissing bugs (*Ben-Yakir, 1987*; *Beard et al., 1992*; *Hypša & Dale, 1997*; *Šorfová, Škeříková & Hypša, 2008*; *Pachebat et al., 2013*), tsetse flies (*Aksoy, 1995*; *Dale & Maudlin, 1999*), bat flies (*Trowbridge, Dittmar & Whiting, 2006*; *Hosokawa et al., 2012*; *Wilkinson et al., 2016*), and louse flies (*Trowbridge, Dittmar & Whiting, 2006*; *Nováková & Hypša, 2007*; *Chrudimský et al., 2012*) have established symbiotic associations with bacteria from different lineages, mostly α-proteobacteria (*Hosokawa et al., 2010*) and γ-proteobacteria (*Aksoy, 1995*; *Hypša & Aksoy, 1997*; *Hypša & Dale, 1997*; *Dale et al., 2006*; *Allen et al., 2007*; *Hypša & Křížek, 2007*; *Nováková & Hypša, 2007*; *Chrudimský et al., 2012*; *Hosokawa et al., 2012*; *Wilkinson et al., 2016*). Obligate symbionts of these blood-sucking hosts are hypothesized to supplement B-vitamins and cofactors missing from their blood diet or present at too low concentration (*Akman et al., 2002*; *Kirkness et al., 2010*; *Rio et al., 2012*; *Nikoh et al., 2014*; *Nováková et al., 2015*; *Boyd et al., 2016*; *Boyd et al., 2017*), but experimental evidence supporting this hypothesis is scarce (*Hosokawa et al., 2010*; *Nikoh et al., 2014*; *Michalkova et al., 2014*; *Snyder & Rio, 2015*). The role played by facultative bacteria in blood-sucking hosts is even less understood, with metabolic or protective function as the two main working hypotheses (*Geiger et al., 2005*; *Geiger et al., 2007*; *Toh et al., 2006*; *Belda et al., 2010*; *Snyder et al., 2010*; *Weiss et al., 2013*).

Due to their medical importance, tsetse flies (Diptera, Glossinidae) belong to the most frequently studied models of such symbioses (*International Glossina Genome Initiative, 2014*). They harbour three different symbiotic bacteria: obligate symbiont *Wigglesworthia glossinidia* which is essential for the host survival (*Pais et al., 2008*), facultative symbiont *Sodalis glossinidius* which was suggested to cooperate with *Wigglesworthia* on thiamine biosynthesis (*Belda et al., 2010*), and reproductive manipulator *Wolbachia* (*Pais et al., 2011*). Considerable amount of information has till now been accumulated on the distribution, genomics and functions of these bacteria (*Akman et al., 2002*; *Toh et al., 2006*; *Rio et al., 2012*; *Balmand et al., 2013*; *Michalkova et al., 2014*; *Snyder & Rio, 2015*). In contrast to our understanding of tsetse fly symbioses, only scarce data are available on the symbioses in its closely related groups. Apart from Glossinidae, the superfamily Hippoboscoidea includes additional three families of obligatory blood-sucking flies, tightly

associated with endosymbionts, namely Nycteribiidae, Streblidae, and Hippoboscidae. Monophyly of Hippoboscoidea has been confirmed by numerous studies (*Nirmala, Hypša & Žurovec, 2001*; *Dittmar et al., 2006*; *Petersen et al., 2007*; *Kutty et al., 2010*), but its inner topology has not been fully resolved. The monophyletic family Glossinidae is considered to be a sister group to the three remaining families together designated as Pupipara (*Petersen et al., 2007*). The two groups associated with bats probably form one branch, where Nycteribiidae seems to be monophyletic while monophyly of Streblidae was not conclusively confirmed (*Dittmar et al., 2006*; *Petersen et al., 2007*; *Kutty et al., 2010*). According to several studies, Hippoboscidae is regarded to be a monophyletic group with not well-resolved exact position in the tree (*Nirmala, Hypša & Žurovec, 2001*; *Dittmar et al., 2006*; *Petersen et al., 2007*). However, louse flies were also shown to be paraphyletic in respect to bat flies (*Dittmar et al., 2006*; *Kutty et al., 2010*).

Nycteribiidae, Streblidae (bat flies), and Hippoboscidae (louse flies) are often associated with *Arsenophonus* bacteria (*Trowbridge, Dittmar & Whiting, 2006*; *Dale et al., 2006*; *Nováková, Hypša & Moran, 2009*; *Morse et al., 2013*; *Duron et al., 2014*). In some cases, these symbionts form clades of obligate lineages coevolving with their hosts, but some of *Arsenophonus* lineages are likely representing loosely associated facultative symbionts spread horizontally across the population (*Nováková, Hypša & Moran, 2009*; *Morse et al., 2013*; *Duron et al., 2014*). Bat flies and louse flies are also commonly infected with *Bartonella* spp. (*Halos et al., 2004*; *Morse et al., 2012b*). *Wolbachia* infection was found in all Hippoboscoidea groups (*Pais et al., 2011*; *Hosokawa et al., 2012*; *Morse et al., 2012a*; *Nováková et al., 2015*). Moreover, several Hippoboscidae species were also found to harbour distinct lineages of *Sodalis*-like bacteria (*Dale et al., 2006*; *Nováková & Hypša, 2007*; *Chrudimský et al., 2012*) likely representing similar facultative-obligatory gradient of symbioses as observed for *Arsenophonus*.

Hippoboscoidea thus represent a group of blood-sucking insects with strikingly dynamic symbioses. Obligate symbionts from *Arsenophonus* and *Sodalis* clades tend to come and go, disrupting the almost flawless host-symbiont co-phylogenies often seen in insect-bacteria systems. However, why are the endosymbiont replacements so common and what keeps the symbiont consortia limited to the specific bacterial clades remains unknown. Tsetse flies as medically important vectors of pathogens are undoubtedly the most studied Hippoboscoidea lineage. However, their low species diversity (22 species), sister relationship to all other clades, and host specificity to mammals, do not allow to draw any general conclusions about the evolution of symbiosis in Hippoboscoidea. To fully understand the symbiotic turn-over, more attention needs to be paid to the neglected Nycteriibidae, Streblidae, and Hippoboscidae lineages. Here, using gene sequencing and draft genome data from all involved partners, we present phylogenies of Hippoboscidae and their symbiont lineages and try to untangle their relationship to the host. In particular, we ask if these are obligate co-evolving lineages, facultative infections, or if they likely represent recent symbiont replacements just re-starting the obligate relationship.

METHODS

Sample collection and DNA isolation

Samples of louse flies were collected in seven countries (South Africa, Papua New Guinea, Ecuador—Galapagos, Vietnam, France, Slovakia, and the Czech Republic; see Table S1 for details), the single sample of bat fly was collected in the Czech Republic. All samples were stored in 96% ethanol at −20 °C. DNA was extracted using the QIAamp DNA Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA quality was verified using the Qubit High Sensitivity Kit (Invitrogen) and 1% agarose gel electrophoresis.

PCR, cloning, and sequencing

All DNA samples were used for amplification of three host genes (COI, 16S rRNA gene, EF) and symbiont screening with 16S rRNA gene primers (Table S2). Ten *Wolbachia* positive samples were used for MLST typing (*cox*A, *fbp*A, *ftsZ*, *gat*B, *hcp*A; see Table S2). PCR reaction was performed under standard conditions using High Fidelity PCR Enzyme Mix (Thermo Scientific, Waltham, MA, USA) and Hot Start Tag DNA Polymerase (Qiagen, Hilden, Germany) according to the manufacturer's protocol. PCR products were analysed using 1% agarose gel electrophoresis and all symbiont 16S rDNA products were cloned into pGEM®–T Easy vector (Promega, Madison, WI, USA) according to the manufacturer's protocol. Inserts from selected colonies were amplified using T7 and SP6 primers or isolated from plasmids using the Jetquick Plasmid Miniprep Spin Kit (Genomed GmbH, Löhne, Germany). Sanger sequencing was performed by an ABI Automatic Sequencer 3730XL (Macrogen Inc., Geumchun-gu-Seoul, Korea) or ABI Prism 310 Sequencer (SEQme, Dobříš, the Czech Republic).

In addition to sequencing, we also included in our analyses genomic data of *Melophagus ovinus* (*Nováková et al., 2015*), *Lipoptena cervi* (*Nováková et al., 2016*), *Ornithomya biloba*, and *Crataerina pallida* (Figshare: https://figshare.com/s/e1488900c5cb62af69ab) as well as their endosymbionts (see Table S1).

Although there is MLST available for *Arsenophonus* bacteria (*Duron, Wilkes & Hurst, 2010*), we were not successful in amplifying these genes.

Alignments and phylogenetic analyses

The assemblies of raw sequences were performed in Geneious v8.1.7 (*Kearse et al., 2012*). Datasets were composed of the assembled sequences, extracted genomic sequences, sequences downloaded from GenBank (see Table S4) or the *Wolbachia* MLST database. The sequences were aligned with Mafft v7.017 (*Katoh, 2002*; *Katoh, Asimenos & Toh, 2009*) implemented in Geneious using an E-INS-i algorithm with default parameters. The alignments were not trimmed as trimming resulted in massive loss of informative position. Phylogenetic analyses were carried out using maximum likelihood (ML) in PhyML v3.0 (*Guindon & Gascuel, 2003*; *Guindon et al., 2009*) and Bayesian inference (BI) in MrBayes v3.1.2 (*Huelsenbeck & Ronquist, 2001*). The $GTR + I + \Gamma$ evolutionary model was selected in jModelTest (*Posada, 2009*) according to the Akaike Information Criterion (AIC). The subtree prunning and regrafting (SPR) tree

search algorithm and 100 bootstrap pseudoreplicates were used in the ML analyses. BI runs were carried out for 10 million generations with default parameters, and Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) was used for convergence and burn-in examination. Phylogenetic trees were visualised and rooted in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and their final graphical adjustments were performed in Inkscape v0.91 (https://inkscape.org/en/).

Host phylogeny was reconstructed using single-gene analyses and a concatenated matrix of three genes (mitochondrial 16S rRNA, mitochondrial cytochrome oxidase I, and nuclear elongation factor). Concatenation of genes was performed in Phyutility 2.2.6 (*Smith & Dunn, 2008*). Phylogenetic trees were inferred for all species from the Hippoboscoidea superfamily, as well as for smaller datasets comprising only Hippoboscidae species. This approach was employed to reveal possible artefacts resulting from missing data and poor taxon-sampling (e.g., short, ~360 bp, sequences of COI available for Streblidae and Nycteribiidae).

Mitochondrial genomes

Problems with reconstruction of host phylogeny based on mitochondrial genes (16S and COI) lead us to assemble mitochondrial genomes of four main louse fly lineages. Contigs of mitochondrial genomes were identified in genomic data of *M. ovinus*, *L. cervi*, *O. biloba*, and *C. pallida* using BLASTn and tBLASTn searches (*Altschul et al., 1990*). Open reading frame identification and preliminary annotations were performed using NCBI BlastSearch in Geneious. For identification of Numts, raw sequences were mapped to mitochondrial data using Bowtie v2.2.3 (*Langmead & Salzberg, 2012*). Web annotation server MITOS (http://mitos.bioinf.uni-leipzig.de/) was used for final annotation of proteins and rRNA/tRNA genes. We selected 15 mitochondrial genes (Table S4) present in all included taxa for phylogenetic inference as described above.

RESULTS

Phylogenetic data

We obtained 138 host sequences: 31 sequences of 16S rRNA of 208–567 bp, 48 sequences of EF of 280–890 bp, and 59 sequences of COI of 299–1,491 bp; and 70 symbiont 16S rRNA sequences of 269–1,210 bp. We also assembled and annotated 4 host mitochondrial genomes of 15,975–16,445 bp. For more details see Table S3 . All raw sequences can be found online in Data S1 (their description is included in Table S6).

Hippoboscidae phylogeny

We reconstructed host phylogeny using three markers: 16S rRNA, EF and COI; as well as mitochondrial genomes. Our analyses of draft genome data revealed that all analysed mitochondrial genomes of louse flies are also present as Numts (nuclear mitochondrial DNA) on the host chromosomes, especially the COI gene often used for phylogenetic analyses. The taxonomically restricted mitochondrial genome matrix verified monophyly of Hippoboscoidea (Fig. S1). Our three-gene dataset yielded only partially resolved and unstable inner Hippoboscoidea phylogeny. Glossinidae and Nycteribiidae formed a well-defined monophyletic groups (only ML analysis of COI did not confirm monophyly

of Nycteribiidae and also did not resolve its relationship to Streblidae), but monophyly of Hippoboscidae and Streblidae was not well supported and different genes/analyses frequently inferred contradictory topologies. Within Hippoboscidae, the position of the Hippoboscinae group and the genus *Ornithoica* were the most problematic (Fig. 1, Figs. S2–S8).

Arsenophonus and *Sodalis* phylogenies

In total, 70 endosymbiont 16S rRNA genes were sequenced in this study and six additional sequences of this gene were mined from our draft genomic data: four of *Arsenophonus*, one of *Sodalis*, and one of *Wolbachia*. Twenty-nine symbionts were identified as members of the genus *Arsenophonus*, 13 symbionts were the most similar to *Sodalis*-allied species, and 28 sequences were of *Wolbachia* origin. Despite cloning, we did not obtain any sequences of *Bartonella* reported to occur in some Hippoboscoidea. Moreover, using only phylogenetic approach, we would not be able to decide whether *Bartonella*-Hippoboscidae interaction is mutualistic or pathogenic, therefore *Bartonella* symbiosis is not in the scope of this manuscript. Putative assignment to the obligate or likely facultative symbiont categories was based on GC content of their 16S rRNA gene and genomic data available (Table S3), branch length, and the phylogenetic analyses.

Phylogenetic analyses of the genus *Arsenophonus* based on 16S rDNA sequences revealed several distinct clades of likely obligate *Arsenophonus* species congruent with their host phylogeny, partially within the Nycteribiidae, Streblidae, and several Hippoboscidae lineages (Fig. 2, Figs. S9 and S10). However, it is important to note that these clades do not form a single monophyletic clade of co-diverging symbionts, but rather several separate lineages. Within Hippoboscidae, the *Arsenophonus* sequences from the Ornithomyini group form a monophyletic clade congruent with Ornithomyini phylogeny except *Arsenophonus* symbiont of *Crataerina* spp. which was probably recently replaced by another *Arsenophonus* bacteria. Other obligate *Arsenophonus* lineages were detected in the genera *Lipoptena*, *Melophagus*, and *Ornithoica*. All other *Arsenophonus* sequences from the Hippoboscidae either represent facultative symbionts or putatively obligate symbioses which are impossible to reliably detect by phylogenetic methods (but see the discussion for *Hippobosca* sp.).

Most of the putatively facultative endosymbionts of the Hippoboscidae typically possess short branches and are also related with the previously described species *Arsenophonus arthropodicus* and *Arsenophonus nasoniae*. Interestingly, both obligate and likely facultative lineages were detected from several species, e.g., *Ornithomya biloba*, *Ornithomya avicularia*, and *Ornithomya fringillina* (Fig. 2). Phylogenetic analyses including symbionts from the genera *Nycterophylia* and *Trichobius* did not clearly place them into the *Arsenophonus* genus. Rather, they likely represent closely related lineages to the *Arsenophonus* clade as their position was unstable and changed with different taxon samplings and methods.

Within *Sodalis*, the phylogenetic reconstruction revealed a putatively obligate endosymbiont from the tribe Olfersini, including the genera *Pseudolynchia* and *Icosta*, and several facultative lineages. However, co-evolution with *Icosta* sp. seems to be imperfect and does not strictly follow the host phylogeny (Fig. 3).

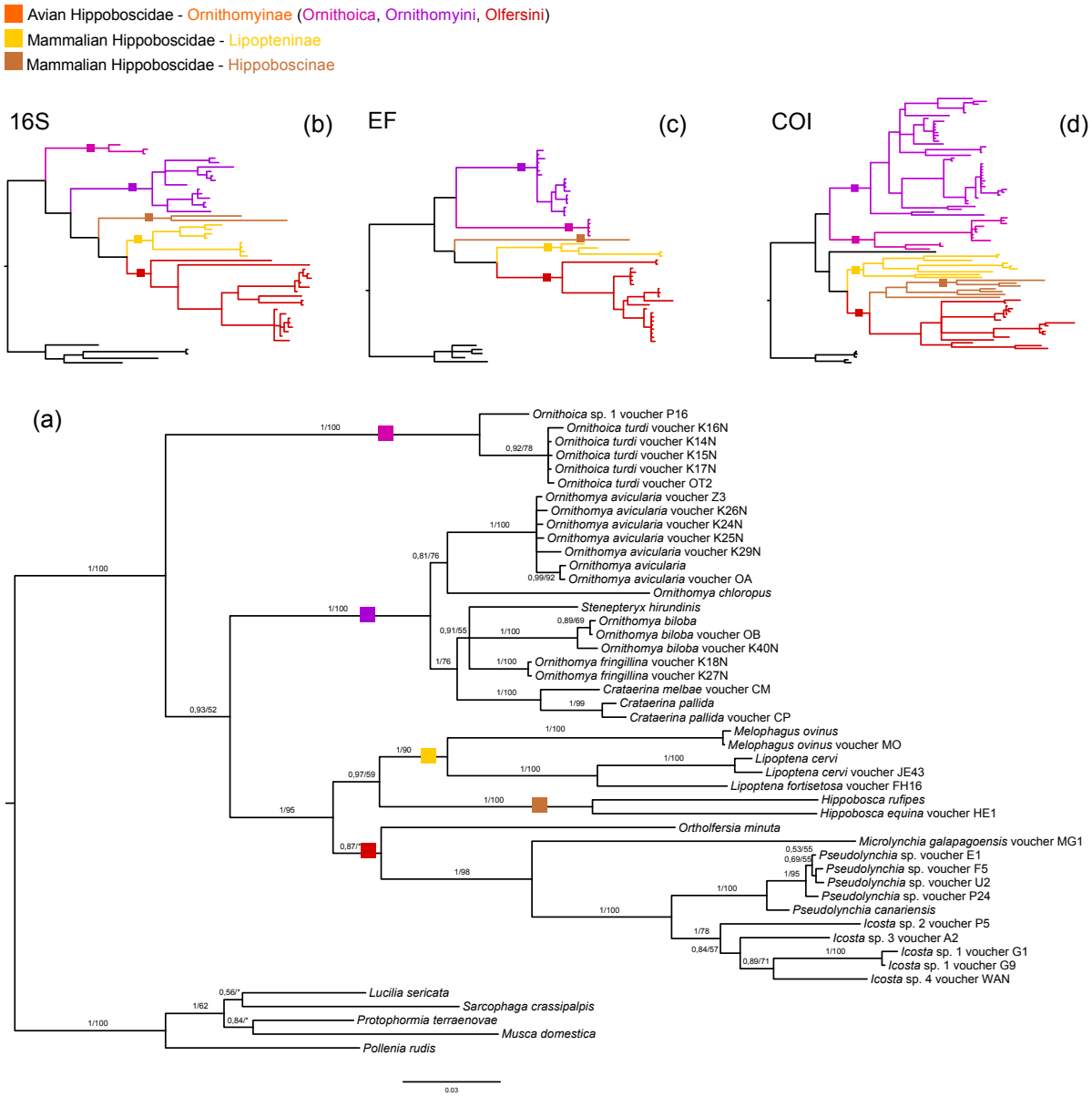


Figure 1 **Host phylogeny.** (A) Host phylogeny derived from concatenation of three genes: 16S rRNA, EF, and COI. The phylogeny was reconstructed by BI analysis. Posterior probabilities and bootstrap support are printed upon branches, respectively (asterisk was used for very low or missing bootstrap branch support). Taxa labelled with voucher are newly sequenced in this study. Genomic COI sequences are labelled with rRNA. Three smaller trees on the top of the figure represent outlines of three separate phylogenetic trees based on BI analyses of 16S rRNA (B), EF (C), and COI (D) genes. Full versions of these phylogenies are included in Figs. S6–S8. Three main families of Hippoboscidae are colour coded: yellow for Lipopteninae (one group), brown for Hippoboscinae (one group), and orange for Ornithomiinae (three groups). Colour squares label branches where are placed main Hippoboscidae groups. This labelling corresponds with labelling of branches at smaller outlines, which are in addition to this highlighted with the same colour. All host trees are included in Figs. S1–S8.

Full-size DOI: 10.7717/peerj.4099/fig-1

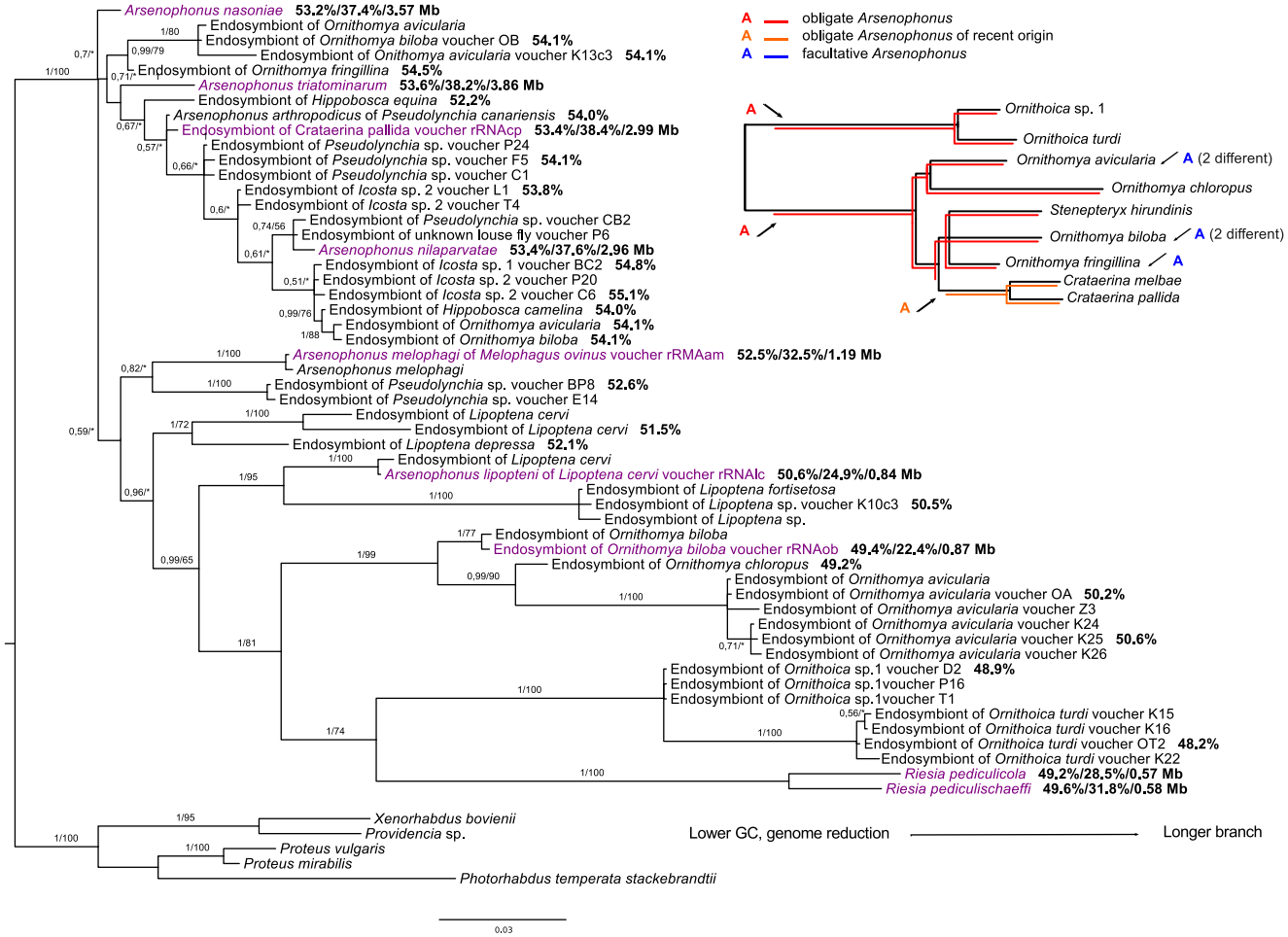


Figure 2 16S rRNA phylogeny of *Arsenophonus* in Hippoboscidae inferred by BI analysis. Posterior probabilities and bootstrap support are printed upon branches, respectively (asterisk was used for very low or missing bootstrap branch support). Taxa labelled with voucher are newly sequenced in this study. Genomic sequences are labelled with rRNA. Taxa in dark purple represent *Arsenophonus* bacteria which genome was sequenced. Numbers behind these taxa correspond to their GC content of 16S rRNA, GC content of genome, and genome size, respectively. Numbers behind other taxa correspond to GC content of their 16S rRNA. The smaller picture on the right side represents host phylogeny to which symbiont phylogeny was compared. Red lineages correspond to obligate symbionts while orange lineage is symbiont of recent origin. The blue A represent likely facultative *Arsenophonus* infection. To achieve this, we also used the information available on *gro*EL gene by *Morse et al. (2013)* and *Duron et al. (2014)*. Phylogenetic reconstructions of *Arsenophonus* of entire Hippoboscoidea and all *Arsenophonus* bacteria are included in Figs. S9 and S10.

Full-size DOI: 10.7717/peerj.4099/fig-2

Wolbachia MLST analysis

In *Wolbachia*, the 16S rDNA sequences were used only for an approximate supergroup determination (Fig. 4A). The MLST analysis was performed with ten selected species (one of them was obtained from genomic data of *O. biloba*; see Table S3). Overall prevalence of *Wolbachia* in louse flies is 54.55%; 30 positive individuals out of 55 diagnosed. The supergroup A was detected from 4 species (4 individuals), the supergroup B from 5 species

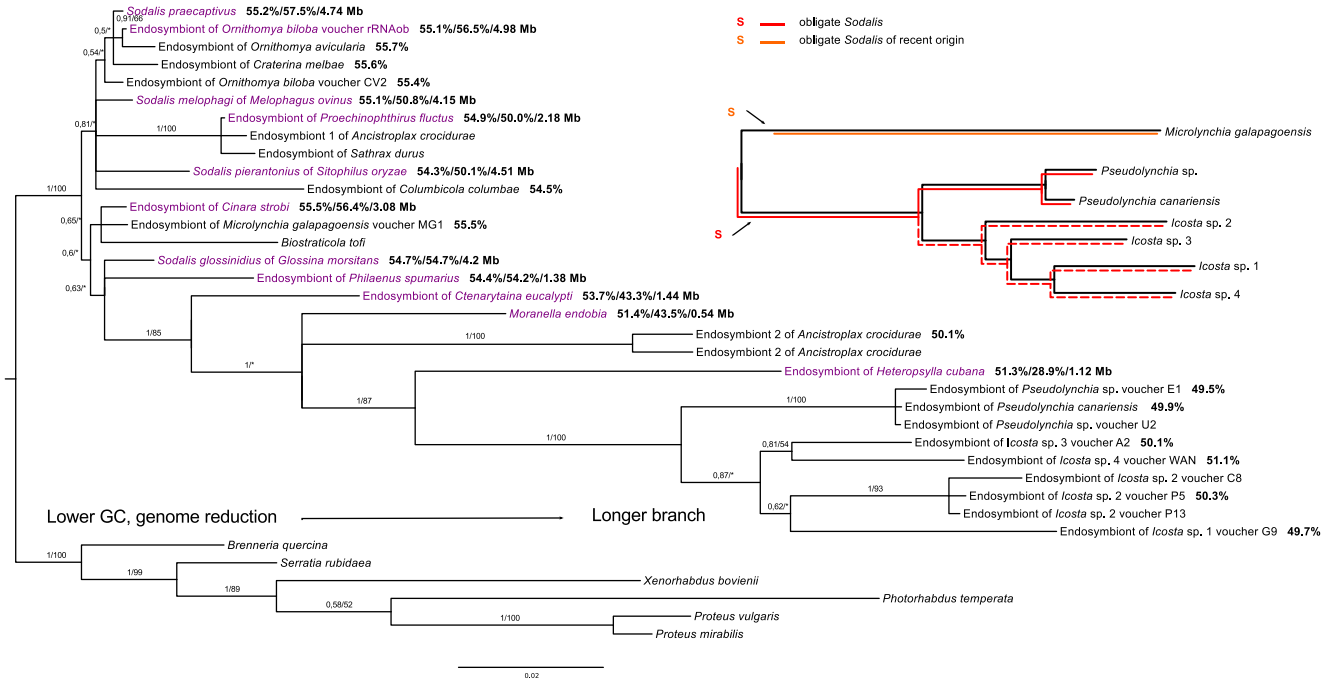


Figure 3 16S rRNA phylogeny of *Sodalis* in Hippoboscidae inferred by BI analysis. Posterior probabilities and bootstrap support are printed upon branches, respectively (asterisk was used for very low or missing bootstrap branch support). Taxa labelled with voucher are newly sequenced in this study. Taxa in dark purple represent *Sodalis*-like bacteria which genome was sequenced. Numbers behind these taxa correspond to their GC content of 16S rRNA, GC content of genome, and genome size, respectively. Numbers behind other taxa correspond to GC content of their 16S rRNA. Red lineages correspond to obligate symbionts while orange lineage is symbiont of recent origin. The red dashed line shows that co-evolution between *Icosta* spp. and their obligate endosymbiont imperfect.

Full-size DOI: 10.7717/peerj.4099/fig-3

(9 individuals), and the supergroup F from 7 species (17 individuals) (Figs. 4A–4B). Additionally, *Nycteribia kolenatii* (one individual) was infected with the supergroup F.

DISCUSSION

Hippoboscidae phylogeny: an unfinished portrait

Although closely related to the medically important tsetse flies, the other hippoboscoids have only rarely been studied and their phylogeny is still unclear. Based on our concatenated matrix, we obtained the topology which to some extent resembles the one presented *Petersen et al. (2007)*, although with slightly different taxon sampling (Fig. 1; Fig. S2). However, our three single-gene datasets implied only poor phylogenetic signal available carried by the hippoboscoid sequences. Therefore, we took an advantage of the four complete mitochondrial genomes reconstructed in this study to test the reliability of the previous phylogenetic reconstructions. The phylogenetic reconstruction based on the mitochondrial matrix correspond to the three-gene concatenated matrix phylogeny suggesting that mitochondrial genomes would be valuable for further phylogenetic analyses of this group (Fig. 1; Fig. S1). According to our results, Glossinidae, Nycteribiidae and

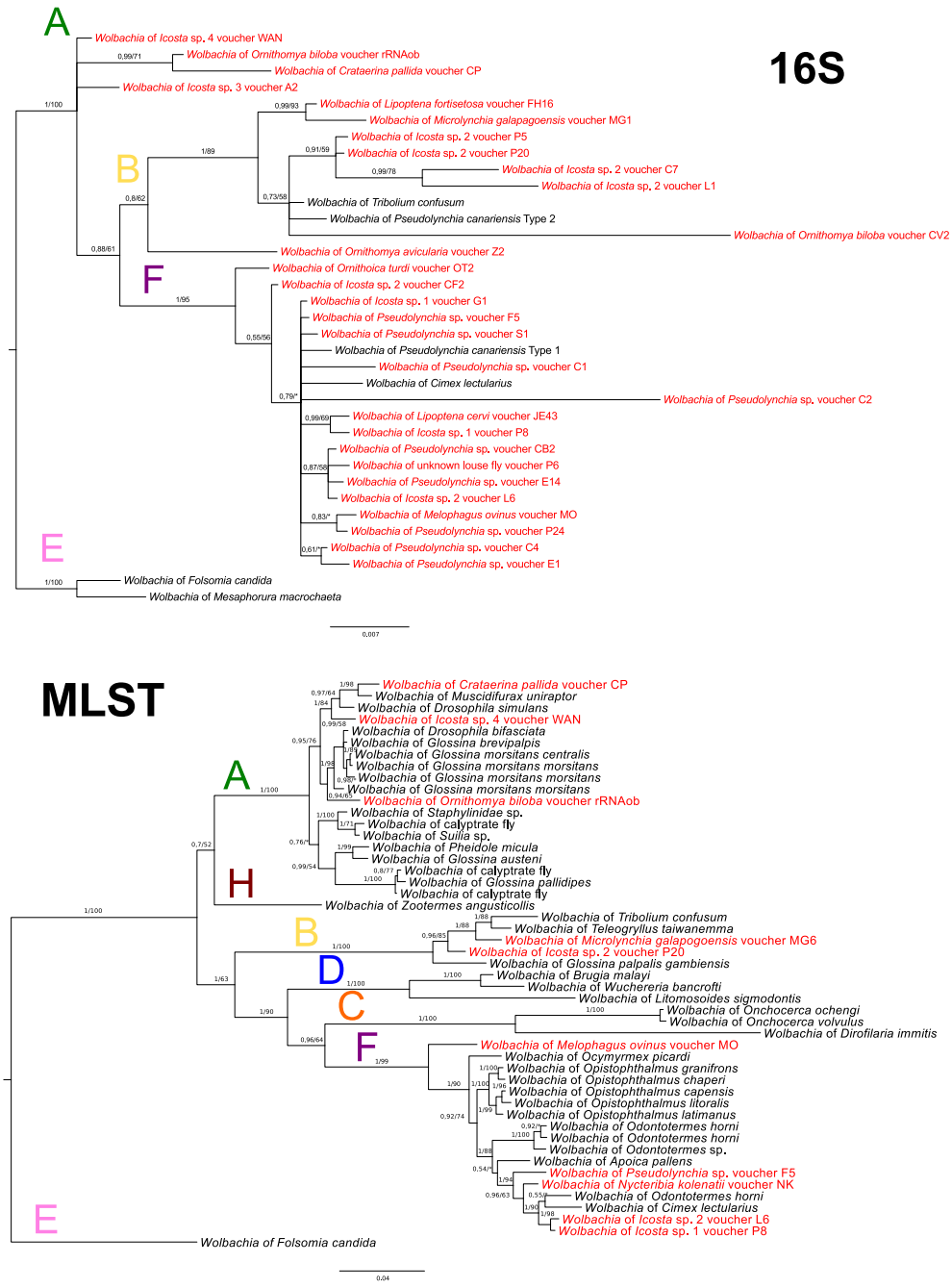


Figure 4 ***Wolbachia*** **phylogeny.** (A) *Wolbachia* phylogeny inferred from 16S rRNA by BI analysis. (B) *Wolbachia* phylogeny inferred from MLST genes by BI analysis. Posterior probabilities and bootstrap support are printed upon branches, respectively (asterisk was used for very low or missing bootstrap branch support). Colour letters upon branches correspond to *Wolbachia* supergroups. Taxa in red represent *Wolbachia* bacteria from Hippoboscidae and Nycteribidae which are newly sequenced in this study. Taxa labelled with # in the 16S tree represent taxa which were used for the MLST analysis. *Wolbachia* from *O. biloba*, which was obtained from genomic data, is labelled with rRNAob. Supergroup E was used for rooting both trees.

Full-size DOI: 10.7717/peerj.4099/fig-4

Hippoboscidae were retained as monophyletic groups, but monophyly of Streblidae was not supported using the complete matrix (Fig. S2). Streblidae lineage appears to be paraphyletic with respect to Nycteribiidae and clusters into two groups, the Old World and the New World species, as previously reported (*Dittmar et al., 2006*; *Kutty et al., 2010*). Within Hippoboscidae, the groups Lipopteninae, Hippoboscinae, Ornithomyini and Olfersini (nomenclature was adopted from *Petersen et al. (2007)*) are well-defined and monophyletic, but their exact relationships are still not clear. The most problematic taxa are Hippoboscinae and also the genus *Ornithoica* with their positions depending on the used genes/analyses (Fig. 1; Figs. S2–S8). A possible explanation for these inconsistencies in the topologies can be a hypothetical rapid radiation from the ancestor of Hippoboscoidea group into main subfamilies of Hippoboscidae leaving in the sequences only very weak phylogenetic signal for this period of Hippoboscidae evolution. The most difficulties in reconstructing Hippoboscoidea phylogeny is caused by missing data (only short sequences of COI are available especially for Nycteribiidae and Streblidae in the GenBank; Fig. S3). Moreover, COI phylogenies are known to be affected by numerous pseudogenes called Numts (*Black IV & Bernhardt, 2009*). The Numts, we found to be common in louse fly genomes, can thus also contribute to the intricacy of presented phylogenies. On the other hand, EF seems to provide plausible phylogenetic information (Fig. S4). The biggest drawback of this marker however lies in the data availability in public databases, restricting an appropriate taxon sampling for the Hippoboscoidea superfamily.

Hidden endosymbiont diversity within the Hippoboscidae family

Among the three most commonly detected Hippoboscidae endosymbionts, attention has been predominantly paid to *Arsenophonus* as the supposedly most common obligate endosymbiont of this group. Our data show that several different lineages of *Arsenophonus* have established the symbiotic lifestyle within Hippoboscidae (Fig. 2). According to our results supported by genomic data, there are at least four lineages of likely obligate endosymbionts: *Arsenophonus* in Ornithomyini (genomes of *Arsenophonus* from *Ornithomya biloba* and *Crataerina pallida* will be published elsewhere), *Arsenophonus* in *Ornithoica* spp., previously described *Arsenophonus melophagi* (*Nováková et al., 2015*) and *Arsenophonus lipopteni* (*Nováková et al., 2016*). All these possess reduced genomes with low GC content as a typical feature of obligate endosymbionts (*McCutcheon & Moran, 2012*). Interestingly, within Ornithomyini, the original obligate *Arsenophonus* endosymbiont of *Crataerina* spp. was recently replaced by another *Arsenophonus* bacterium with ongoing genome reduction (Figshare: https://figshare.com/s/e1488900c5cb62af69ab). Apart from these potentially obligate lineages, there are other hippoboscid associated *Arsenophonus* bacteria distributed in the phylogenetic tree among *Arsenophonus* endosymbionts with likely facultative or free-living lifestyle (Fig. S10). This pattern suggests *Arsenophonus* is likely being repeatedly acquired from the environment. It has been hypothesized that obligate endosymbionts often evolve from facultative symbionts which are no longer capable of horizontal transmission between the hosts (*Moran, McCutcheon & Nakabachi, 2008*). Due to their recent change of lifestyle, endosymbionts with an ongoing genome reduction in many ways resemble facultative symbionts, e.g., their positions in phylogenetic

trees are not stable and differ with the analysis method and taxon sampling (Fig. 2, Figs. S9 and S10). Such nascent stage of endosymbiosis was indicated for the obligate *Arsenophonus* endosymbiont of *C. pallida* (Figshare: https://figshare.com/s/e1488900c5cb62af69ab) and similar results can be expected for *Arsenophonus* endosymbionts of *Hippobosca* species.

Within bat flies, we found obligate *Arsenophonus* lineages in both Nycteribiidae and Streblidae as well as several presumably facultative *Arsenophonus* infections in both groups (Figs. S9 and S10). Similar results were reported in several previous studies (*Morse et al., 2013*; *Duron et al., 2014*; *Wilkinson et al., 2016*). Members of the *Arsenophonus* clade were also reported from Nycterophyliinae and Trichobiinae (Streblidae) (*Morse et al., 2012a*) and *Cyclopodia dubia* (Nycteribiidae) (*Wilkinson et al., 2016*). However, our results do not support their placement within the clade, as these sequences were attracted by the long branches in the ML analyses. The endosymbiont of Nycterophyliinae and Trichobiinae probably represents an ancient lineage closely related to *Arsenophonus* clade (Fig. S9) while the endosymbiont of *Cyclopodia dubia* is more likely related with *Pectobacterium* spp.; therefore, we excluded this bacterium from our further analyses. These findings indicate that bat flies established the endosymbiotic lifestyle several times independently with at least three bacterial genera.

In contrast to *Arsenophonus*, only a few studies reported *Sodalis*-like endosymbiotic bacteria from Hippoboscidae (*Nováková & Hypša, 2007*; *Chrudimský et al., 2012*; *Nováková et al., 2015*). *Dale et al. (2006)* detected a putative obligate endosymbiont from *Pseudolynchia canariensis* which was suggested to represent *Sodalis* bacterium. We detected this symbiont in several members of the Olfersini group and according to our results, it is obligate *Sodalis*-like endosymbiont forming a monophyletic clade, but its congruence with the Olfersini phylogeny is somewhat imperfect (Fig. 3). This incongruence might be a consequence of phylogenetic artefacts likely affecting long branches of *Sodalis* symbionts from *Icosta*. Similar to *Arsenophonus*, *Sodalis* bacteria also establish possible facultative associations, e.g., with *Melophagus ovinus* (*Chrudimský et al., 2012*; *Nováková et al., 2015*), *Ornithomya avicularia* (*Chrudimský et al., 2012*) or *Ornithomya biloba* (this study). *Sodalis* endosymbiont from *Crataerina melbae* was suggested to be obligate (*Nováková & Hypša, 2007*), but our study did not support this hypothesis since it clusters with free-living *Sodalis praecaptivus*. Interestingly, *Sodalis* endosymbiont of *Microlynchia galapagoensis* was inferred to be closely related to *Sodalis*-like co-symbiont of *Cinara cedri*, which underwent rapid genome deterioration after a replacement of former co-symbiont (*Meseguer et al., 2017*). These results suggest that there are several loosely associated lineages of *Sodalis* bacteria in louse flies. On one hand, the endosymbiont of *Microlynchia galapagoensis* probably represents a separate (or ancient) *Sodalis* infection, but on the other hand, other *Sodalis* infections seem to be repeatedly acquired from the environment as implied by their relationship to e.g., *Sodalis praecaptivus* (*Clayton et al., 2012*) (Fig. 3).

Coinfections of obligate and facultative *Arsenophonus* strains in Hippoboscidae (or potentially *Sodalis* in Olfersini) are extremely difficult to recognize using only PCR-acquired 16S rRNA gene. Facultative endosymbionts retain several copies of this gene and thus their 16S rRNA tend to be amplified more likely in PCR than from reduced obligate endosymbionts due to its higher copy number and lower frequency of mutations in primer

binding sites. Even though there is a MLST available for *Arsenophonus* bacteria (*Duron, Wilkes & Hurst, 2010*), it was shown that it is effective only partially (*Duron et al., 2014*). Since our data are probably also influenced by this setback, we do not speculate which of the detected potentially facultative *Arsenophonus* lineages represent source of 'ancestors' for several distinct obligate lineages or which of them were involved in the recent replacement scenario. However, the replacement/independent-origin scenario is well illustrated by endosymbionts from Olfersini (Figs. 2 and 3).

To complement the picture of Hippoboscidae endosymbiosis, we also reconstructed *Wolbachia* evolution. We found three different supergroups: A, B and F (see Table S3). Apparently, there is no coevolution between *Wolbachia* and Hippoboscidae hosts suggesting horizontal transmission between species (Figs. 4A–4B) as common for this bacterium (*Schilthuizen & Stouthamer, 1997*; *Gerth et al., 2014*). Since *Wolbachia* seems to be one of the most common donors of genes horizontally transferred to insect genomes, including tsetse flies (*Husník et al., 2013*; *Brelsfoard et al., 2014*; *Sloan et al., 2014*), we cannot rule out that some of *Wolbachia* sequences detected in this study represent HGT insertions into the respective host genomes. The biological role of *Wolbachia* in Hippoboscidae was never examined in spite of its relatively high prevalence in this host group (55%). The F supergroup was detected as the most frequent lineage in Hippoboscidae which is congruent with its common presence in blood-sucking insects such as Streblidae (*Morse et al., 2012a*), Nycteribiidae (*Hosokawa et al., 2012*), Amblycera (*Covacin & Barker, 2007*), and Cimicidae (*Hosokawa et al., 2010*; *Nikoh et al., 2014*).

Besides the three main Hippoboscidae symbionts we paid attention to, *Bartonella* spp. that are also widespread among louse flies and bat flies. The infection seems to be fixed only in *Melophagus ovinus* suggesting a mutualistic relationship (*Halos et al., 2004*), but additional functional data are needed to confirm this hypothesis (*Nováková et al., 2015*). Nevertheless, deer ked and sheep ked are also suspected of vectoring bartonellosis (*Maggi et al., 2009*; *De Bruin et al., 2015*). According to the recent findings, *Bartonella* spp. used to be originally gut symbionts which adapted to pathogenicity (*Segers et al., 2016*; *Neuvonen et al., 2016*).

What is behind dynamics of Hippoboscidae-symbiont associations?

According to our results, symbiosis in the Hippoboscidae group is very dynamic and influenced by frequent symbiont replacements. *Arsenophonus* and *Sodalis* infections seem to be the best resources for endosymbiotic counterparts, but it remains unclear why just these two genera. Both are endowed with several features of free-living/pathogenic bacteria enabling them to enter new host which can be crucial in establishing novel symbiotic association. *Sodalis glossinidius* possesses modified outer membrane protein (OmpA) which is playing an important role in the interaction with the host immune system (*Weiss et al., 2008*; *Weiss, Maltz & Aksoy, 2012*). Both *Sodalis* and *Arsenophonus* bacteria retain genes for the type III secretion system (*Dale et al., 2001*; *Wilkes et al., 2010*; *Chrudimský et al., 2012*; *Oakeson et al., 2014*) allowing pathogenic bacteria to invade eukaryotic cells. Moreover, several strains of these bacteria are cultivable under laboratory conditions (*Hypša & Dale, 1997*; *Dale & Maudlin, 1999*; *Dale et al., 2006*; *Darby et al., 2010*; *Chrudimský et al., 2012*;

Chari et al., 2015) suggesting that they should be able to survive horizontal transmission. For instance, *Arsenophonus nasoniae* is able to spread by horizontal transfer between species (*Duron, Wilkes & Hurst, 2010*), while *Sodalis*-allied bacteria have several times successfully replaced ancient symbionts (*Conord et al., 2008*; *Koga et al., 2013*; *Meseguer et al., 2017*).

Whereas the facultative endosymbionts of Hippoboscoidea are widespread in numerous types of tissues such as milk glands, bacteriome, haemolymph, gut, fat body, and reproductive organs (*Dale & Maudlin, 1999*; *Dale et al., 2006*; *Balmand et al., 2013*; *Nováková et al., 2015*), the obligate endosymbionts are restricted to the bacteriome and milk glands (*Aksoy, 1995*; *Attardo et al., 2008*; *Balmand et al., 2013*; *Morse et al., 2013*; *Nováková et al., 2015*). Entering the milk glands ensures vertical transmission of facultative endosymbiont to progeny and better establishment of the infection. Vertical transmission also enables the endosymbiont to hitch-hike with the obligate endosymbiont and because the obligate endosymbiont is inevitably degenerating (*Moran, 1996*; *Wernegreen, 2002*), the new co-symbiont can eventually replace it if needed. For instance, *Sodalis melophagi* was shown to appear in both milk glands and bacteriome and to code for the same full set of B-vitamin pathways (including in addition the thiamine pathway) as the obligate endosymbiont *Arsenophonus melophagi* (*Nováková et al., 2015*). This suggests that it could be potentially capable of shifting from facultative to obligatory lifestyle and replace the *Arsenophonus melophagi* endosymbiont.

We suggest that the complex taxonomic structure of the symbiosis in Hippoboscoidea can be result of multiple replacements, similar to that already suggested for the evolution of symbiosis in *Columbicola* lice (*Smith et al., 2013*) or mealybugs (*Husník & McCutcheon, 2016*). Based on the arrangement of the current symbioses in various species of Pupipara, the ancestral endosymbiont was likely either an *Arsenophonus* or *Sodalis* bacterium (given our finding of the potential obligate *Sodalis* lineage in Olfersini). In the course of Pupipara evolution and speciation, this symbiont was repeatedly replaced by different *Arsenophonus* (or *Sodalis* in Olfersini if not ancestral) lineages, as indicated by the lack of phylogenetic congruence and differences in genome reduction, gene order, and GC content in separate *Arsenophonus* lineages (*Nováková et al., 2015*; *Nováková et al., 2016*; Figshare: https://figshare.com/s/e1488900c5cb62af69ab). This genomic diversity across the *Arsenophonus* bacteria from distinct Hippoboscidae thus likely reflects their different age correlating with the level of genome reduction in symbiotic bacteria.

CONCLUSIONS

Despite the considerable ecological and geographical variability, the Hippoboscoidea families surprisingly share some aspects of their association with symbiotic bacteria. Particularly, they show high affinity to two bacterial genera, *Arsenophonus* and *Sodalis*. This affinity is not only reflected by frequent occurrence of the bacteria but mainly by their multiple independent acquisitions. Comparisons between the hippoboscid and bacterial phylogenies indicate several independent origins of the symbiosis, although more precise evolutionary reconstruction is still hampered by the uncertainties in hippoboscid phylogenies.

ACKNOWLEDGEMENTS

Ali Halajian thanks Prof. Wilmien J. Luus-Powell (Biodiversity Research Chair, University of Limpopo) for support and Prof. Derek Engelbrecht for help with bird catching and identifications.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Grant Agency of the Czech Republic (grant 13-01878S to Václav Hypša). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Grant Agency of the Czech Republic: 13-01878S.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Eva Šochová performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Filip Husník and Václav Hypša conceived and designed the experiments, reviewed drafts of the paper.
- Eva Nováková performed the experiments, reviewed drafts of the paper.
- Ali Halajian collected samples of African louse flies.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Raw sequences are included in Data S1. Matrices and genomic data of four bacteria used for our analyses are deposited at Figshare (https://figshare.com/s/e1488900c5cb62af69ab).

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4099#supplemental-information.

REFERENCES

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics* **32**:402–407 DOI 10.1038/ng986.

Aksoy S. 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *International Journal of Systematic Bacteriology* **45**:848–851 DOI 10.1099/00207713-45-4-848.

Allen JM, Burleigh JG, Light JE, Reed DL. 2016. Effects of 16S rDNA sampling on estimates of the number of endosymbiont lineages in sucking lice. *PeerJ* **4**:e2187 DOI 10.7717/peerj.2187.

Allen JM, Reed DL, Perotti MA, Braig HR. 2007. Evolutionary relationships of "*Candidatus* Riesia spp.", endosymbiotic enterobacteriaceae living within hematophagous primate lice. *Applied and Environmental Microbiology* **73**:1659–1664 DOI 10.1128/AEM.01877-06.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410 DOI 10.1016/S0022-2836(05)80360-2.

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. *Journal of Insect Physiology* **54**:1236–1242 DOI 10.1016/j.jinsphys.2008.06.008.

Balmand S, Lohs C, Aksoy S, Heddi A. 2013. Tissue distribution and transmission routes for the tsetse fly endosymbionts. *Journal of Invertebrate Pathology* **112**:S116–S122 DOI 10.1016/j.jip.2012.04.002.

Beard CB, Mason PW, Aksoy S, Tesh RB, Richards FF. 1992. Transformation of an insect symbiont and expression of a foreign gene in the Chagas' disease vector *Rhodnius prolixus*. *The American Journal of Tropical Medicine and Hygiene* **46**:195–200 DOI 10.4269/ajtmh.1992.46.195.

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

Ben-Yakir D. 1987. Growth retardation of *Rhodnius prolixus* symbionts by immunizing host against *Nocardia* (*Rhodococcus*) *rhodnii*. *Journal of Insect Physiology* **33**:379–383 DOI 10.1016/0022-1910(87)90015-1.

Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proceedings of the National Academy of Sciences of the United States of America* **112**:10169–10176 DOI 10.1073/pnas.1421388112.

Black IV WC, Bernhardt SA. 2009. Abundant nuclear copies of mitochondrial origin (NUMTs) in the *Aedes aegypti* genome. *Insect Molecular Biology* **18**:705–713 DOI 10.1111/j.1365-2583.2009.00925.x.

Boyd BM, Allen JM, Koga R, Fukatsu T, Sweet AD, Johnson KP, Reed DL. 2016. Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Applied and Environmental Microbiology* **82**:3185–3197 DOI 10.1128/AEM.00282-16.

Boyd BM, Allen JM, Nguyen N-P, Vachaspati P, Quicksall ZS, Warnow T, Mugisha L, Johnson KP, Reed DL. 2017. Primates, lice and bacteria: speciation and genome evolution in the symbionts of hominid lice. *Molecular Biology and Evolution* **34**:1743–1757 DOI 10.1093/molbev/msx117.

Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U, Doudoumis V, Scolari F, Benoit JB, Swain M, Takac P, Malacrida AR, Bourtzis K, Aksoy S. 2014. Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. *PLOS Neglected Tropical Diseases* **8**:e2728 DOI 10.1371/journal.pntd.0002728.

Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, Dale C. 2015. Phenotypic characterization of *Sodalis* praecaptivus sp. nov., a close non-insect-associated member of the *Sodalis*-allied lineage of insect endosymbionts. *International Journal of Systematic and Evolutionary Microbiology* **65**:1400–1405 DOI 10.1099/ijs.0.000091.

Chrudimský T, Husník F, Nováková E, Hypša V. 2012. *Candidatus Sodalis* melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius*. *PLOS ONE* **7**:e40354 DOI 10.1371/journal.pone.0040354.

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

Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G, Heddi A. 2008. Long-term evolutionary stability of bacterial endosymbiosis in curculionoidea: additional evidence of symbiont replacement in the dryophthoridae family. *Molecular Biology and Evolution* **25**:859–868 DOI 10.1093/molbev/msn027.

Covacin C, Barker SC. 2007. Supergroup F *Wolbachia* bacteria parasitise lice (Insecta: Phthiraptera). *Parasitology Research* **100**:479–485 DOI 10.1007/s00436-006-0309-6.

Dale C, Beeton M, Harbison C, Jones T, Pontes M. 2006. Isolation, pure culture, and characterization of "*Candidatus* Arsenophonus arthropodicus", an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. *Applied and Environmental Microbiology* **72**:2997–3004 DOI 10.1128/AEM.72.4.2997-3004.2006.

Dale C, Maudlin I. 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *International Journal of Systematic Bacteriology* **49(Pt 1)**:267–275 DOI 10.1099/00207713-49-1-267.

Dale C, Young SA, Haydon DT, Welburn SC. 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *Proceedings of the National Academy of Sciences of the United States of America* **98**:1883–1888 DOI 10.1073/pnas.98.4.1883.

Darby AC, Choi J-H, Wilkes T, Hughes MA, Werren JH, Hurst GDD, Colbourne JK. 2010. Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*. *Insect Molecular Biology* **19(Suppl 1)**:75–89 DOI 10.1111/j.1365-2583.2009.00950.x.

De Bruin A, Van Leeuwen AD, Jahfari S, Takken W, Földvári M, Dremmel L, Sprong H, Földvári G. 2015. Vertical transmission of *Bartonella schoenbuchensis* in *Lipoptena cervi*. *Parasites & Vectors* **8**:176–181 DOI 10.1186/s13071-015-0764-y.

Dittmar K, Porter ML, Murray S, Whiting MF. 2006. Molecular phylogenetic analysis of nycteribiid and streblid bat flies (Diptera: Brachycera, Calyptratae): implications for host associations and phylogeographic origins. *Molecular Phylogenetics and Evolution* **38**:155–170 DOI 10.1016/j.ympev.2005.06.008.

Duron O, Schneppat UE, Berthomieu A, Goodman SM, Droz B, Paupy C, Nkoghe JO, Rahola N, Tortosa P. 2014. Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. *Molecular Ecology* **23**:2105–2117 DOI 10.1111/mec.12704.

Duron O, Wilkes TE, Hurst GDD. 2010. Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecology Letters* **13**:1139–1148 DOI 10.1111/j.1461-0248.2010.01502.x.

Fukatsu T, Hosokawa T, Koga R, Nikoh N, Kato T, Hayama S, Takefushi H, Tanaka I. 2009. Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: distinct endosymbiont origins in anthropoid primate lice and the old world monkey louse. *Applied and Environmental Microbiology* **75**:3796–3799 DOI 10.1128/AEM.00226-09.

Geiger A, Ravel S, Frutos R, Cuny G. 2005. *Sodalis glossinidius* (Enterobacteriaceae) and vectorial competence of *Glossina palpalis gambiensis* and *Glossina morsitans morsitans* for *Trypanosoma congolense* savannah type. *Current Microbiology* **51**:35–40 DOI 10.1007/s00284-005-4525-6.

Geiger A, Ravel S, Mateille T, Janelle J, Patrel D, Cuny G, Frutos R. 2007. Vector competence of *Glossina palpalis gambiensis* for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*. *Molecular Biology and Evolution* **24**:102–109 DOI 10.1093/molbev/msl135.

Gerth M, Gansauge M-T, Weigert A, Bleidorn C. 2014. Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. *Nature Communications* **5**:5117–5123 DOI 10.1038/ncomms6117.

Guindon S, Delsuc F, Dufayard J-F, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. In: Posada D, ed. *Bioinformatics for DNA sequence analysis*. Totowa: Humana Press, 113–137.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**:696–704 DOI 10.1080/10635150390235520.

Halos L, Jamal T, Maillard R, Girard B, Guillot J, Chomel B, Vayssier-Taussat M, Boulouis H-J. 2004. Role of Hippoboscidae flies as potential vectors of *Bartonella* spp. infecting wild and domestic ruminants. *Applied and Environmental Microbiology* **70**:6302–6305 DOI 10.1128/AEM.70.10.6302-6305.2004.

Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proceedings of the National Academy of Sciences of the United States of America* **107**:769–774 DOI 10.1073/pnas.0911476107.

Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng X-Y, Fukatsu T. 2012. Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *The ISME Journal* **6**:577–587 DOI 10.1038/ismej.2011.125.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755 DOI 10.1093/bioinformatics/17.8.754.

Husník F, McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **113**:E5416–E5424 DOI 10.1073/pnas.1603910113.

Husník F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC, Von Dohlen CD, Fukatsu T, McCutcheon JP. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *CELL* **153**:1567–1578 DOI 10.1016/j.cell.2013.05.040.

Hypša V, Aksoy S. 1997. Phylogenetic characterization of two transovarially transmitted endosymbionts of the bedbug *Cimex lectularius* (Heteroptera: Cimicidae). *Insect Molecular Biology* **6**:301–304 DOI 10.1046/j.1365-2583.1997.00178.x.

Hypša V, Dale C. 1997. *In vitro* culture and phylogenetic analysis of "*Candidatus Arsenophonus* triatominarum", an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *International Journal of Systematic Bacteriology* **47**:1140–1144 DOI 10.1099/00207713-47-4-1140.

Hypša V, Křížek J. 2007. Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). *Microbial Ecology* **54**:242–251 DOI 10.1007/s00248-006-9194-x.

International Glossina Genome Initiative IGG. 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science* **344**:380–386 DOI 10.1126/science.1249656.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**:3059–3066 DOI 10.1093/nar/gkf436.

Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. In: Posada D, ed. *Bioinformatics for DNA sequence analysis*. Totowa: Humana Press, 39–64.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649 DOI 10.1093/bioinformatics/bts199.

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li H-M, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP,

Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner RV, Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America* **107**:12168–12173 DOI 10.1073/pnas.1003379107.

Koga R, Bennett GM, Cryan JR, Moran NA. 2013. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environmental Microbiology* **15**:2073–2081 DOI 10.1111/1462-2920.12121.

Koga R, Tsuchida T, Sakurai M, Fukatsu T. 2007. Selective elimination of aphid endosymbionts: effects of antibiotic dose and host genotype, and fitness consequences. *FEMS Microbiology Ecology* **60**:229–239 DOI 10.1111/j.1574-6941.2007.00284.x.

Kutty SN, Pape T, Wiegmann BM, Meier R. 2010. Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Systematic Entomology* **35**:614–635 DOI 10.1111/j.1365-3113.2010.00536.x.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359 DOI 10.1038/nmeth.1923.

Maggi RG, Kosoy M, Mintzer M, Breitschwerdt EB. 2009. Isolation of *Candidatus* Bartonella melophagi from human blood. *Emerging Infectious Diseases* **15**:66–68 DOI 10.3201/eid1501.081080.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nature Reviews. Microbiology* **10**:13–26 DOI 10.1038/nrmicro2670.

Meseguer AS, Manzano-Marín A, Coeur d'Acier A, Clamens A-L, Godefroid M, Jousselin E. 2017. *Buchnera* has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Molecular Ecology* **26**:2363–2378 DOI 10.1111/mec.13910.

Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. 2014. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Applied and Environmental Microbiology* **80**:5844–5853 DOI 10.1128/AEM.01150-14.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **93**:2873–2878 DOI 10.1073/pnas.93.7.2873.

Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* **42**:165–190 DOI 10.1146/annurev.genet.41.110306.130119.

Morse SF, Bush SE, Patterson BD, Dick CW, Gruwell ME, Dittmar K. 2013. Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). *Applied and Environmental Microbiology* **79**:2952–2961 DOI 10.1128/AEM.03814-12.

Morse SF, Dick CW, Patterson BD, Dittmar K. 2012a. Some like it hot: evolution and ecology of novel endosymbionts in bat flies of cave-roosting bats (Hippoboscoidea, Nycterophiliinae). *Applied and Environmental Microbiology* **78**:8639–8649 DOI 10.1128/AEM.02455-12.

Morse SF, Olival KJ, Kosoy M, Billeter S, Patterson BD, Dick CW, Dittmar K. 2012b. Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). *Infection, Genetics and Evolution* **12**:1717–1723 DOI 10.1016/j.meegid.2012.06.009.

Neuvonen M-M, Tamarit D, Näslund K, Liebig J, Feldhaar H, Moran NA, Guy L, Andersson SGE. 2016. The genome of Rhizobiales bacteria in predatory ants reveals urease gene functions but no genes for nitrogen fixation. *Scientific Reports* **6**:39197 DOI 10.1038/srep39197.

Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences of the United States of America* **111**:10257–10262 DOI 10.1073/pnas.1409284111.

Nirmala X, Hypša V, Žurovec M. 2001. Molecular phylogeny of Calyptratae (Diptera: Brachycera): the evolution of 18S and 16S ribosomal rDNAs in higher dipterans and their use in phylogenetic inference. *Insect Molecular Biology* **10**:475–485.

Nováková E, Husník F, Šochová E, Hypša V. 2015. *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and Environmental Microbiology* **81**:6189–6199 DOI 10.1128/AEM.01487-15.

Nováková E, Hypša V. 2007. A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS Microbiology Letters* **269**:131–135 DOI 10.1111/j.1574-6968.2006.00620.x.

Nováková E, Hypša V, Moran NA. 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiology* **9**:143 DOI 10.1186/1471-2180-9-143.

Nováková E, Hypša V, Nguyen P, Husník F, Darby AC. 2016. Genome sequence of *Candidatus Arsenophonus* lipopteni, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *StandArds in Genomic Sciences* **11**:1–7 DOI 10.1186/s40793-016-0195-1.

Oakeson KF, Gil R, Clayton AL, Dunn DM, Von Niederhausern AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, Vallier A, Jackson DG, Latorre A, Weiss RB, Heddi A, Moya A, Dale C. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biology and Evolution* **6**:76–93 DOI 10.1093/gbe/evt210.

Pachebat JA, Van Keulen G, Whitten MMA, Girdwood S, Del Sol R, Dyson PJ, Facey PD. 2013. Draft genome sequence of *Rhodococcus rhodnii* strain LMG5362,

a symbiont of *Rhodnius prolixus* (Hemiptera, Reduviidae, Triatominae), the principle vector of *Trypanosoma cruzi*. *Genome Announcements* **1**:e00329-13 DOI 10.1128/genomeA.00329-13.

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

Pais R, Lohs C, Wu Y, Wang J, Aksoy S. 2008. The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology* **74**:5965–5974 DOI 10.1128/AEM.00741-08.

Petersen FT, Meier R, Kutty SN, Wiegmann BM. 2007. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. *Molecular Phylogenetics and Evolution* **45**:111–122 DOI 10.1016/j.ympev.2007.04.023.

Posada D. 2009. Selection of models of DNA evolution with jModelTest. In: Posada D, ed. *Bioinformatics for DNA sequence analysis*. Totowa: Humana Press, 93–112.

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

Schilthuizen M, Stouthamer R. 1997. Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps. *Proceedings. Biological Sciences/the Royal Society* **264**:361–366 DOI 10.1098/rspb.1997.0052.

Segers FH, Kešnerová L, Kosoy M, Engel P. 2016. Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *The ISME Journal* **11(5)**:1232–1244 DOI 10.1038/ismej.2016.201.

Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution* **31**:857–871 DOI 10.1093/molbev/msu004.

Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**:715–716 DOI 10.1093/bioinformatics/btm619.

Smith WA, Oakeson KF, Johnson KP, Reed DL, Carter T, Smith KL, Koga R, Fukatsu T, Clayton DH, Dale C. 2013. Phylogenetic analysis of symbionts in feather-feeding lice of the genus Columbicola: evidence for repeated symbiont replacements. *BMC Evolutionary Biology* **13**:1 DOI 10.1186/1471-2148-13-109.

Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM. 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proceedings of the Royal Society B: Biological Sciences* **277**:2389–2397 DOI 10.1098/rspb.2010.0364.

Snyder AK, Rio RVM. 2015. "*Wigglesworthia morsitans*" folate (Vitamin B9) biosynthesis contributes to tsetse host fitness. *Applied and Environmental Microbiology* **81**:5375–5386 DOI 10.1128/AEM.00553-15.

Sudakaran S, Kost C, Kaltenpoth M. 2017. Symbiont acquisition and replacement as a source of ecological innovation. *Trends in Microbiology* **25**:375–390 DOI 10.1016/j.tim.2017.02.014.

Toh H, Weiss BL, Perkin S a H, Yamashita A, Oshima K, Hattori M, Aksoy S. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Research* **16**:149–156 DOI 10.1101/gr.4106106.

Trowbridge RE, Dittmar K, Whiting MF. 2006. Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *Journal of Invertebrate Pathology* **91**:64–68 DOI 10.1016/j.jip.2005.08.009.

Šorfová P, Škeříková A, Hypša V. 2008. An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of *Arsenophonus triatominarum*. *Systematic and Applied Microbiology* **31**:88–100 DOI 10.1016/j.syapm.2008.02.004.

Weiss BL, Maltz MA, Aksoy S. 2012. Obligate symbionts activate immune system development in the tsetse fly. *Journal of Immunology* **188**:3395–3403 DOI 10.4049/jimmunol.1103691.

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

Weiss BL, Wu Y, Schwank JJ, Tolwinski NS, Aksoy S. 2008. An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. *Proceedings of the National Academy of Sciences of the United States of America* **105**:15088–15093 DOI 10.1073/pnas.0805666105.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nature Reviews. Genetics* **3**:850–861 DOI 10.1038/nrg931.

Wilkes TE, Darby AC, Choi J-H, Colbourne JK, Werren JH, Hurst GDD. 2010. The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. *Insect Molecular Biology* **19(Suppl 1)**:59–73 DOI 10.1111/j.1365-2583.2009.00963.x.

Wilkinson DA, Duron O, Cordonin C, Gomard Y, Ramasindrazana B, Mavingui P, Goodman SM, Tortosa P. 2016. The bacteriome of bat flies (Nycteribiidae) from the Malagasy region: a community shaped by host ecology, bacterial transmission mode, and host-vector specificity. *Applied and Environmental Microbiology* **82**:1778–1788 DOI 10.1128/AEM.03505-15.

Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* **32**:723–735 DOI 10.1111/j.1574-6976.2008.00123.x.

Paper 10:

Říhová, J., **Nováková, E**., Husník, F., & Hypša, V. (2017). Legionella becoming a mutualist: adaptive processes shaping the genome of symbiont in the louse Polyplax serrata. *Genome biology and evolution*, *9*(11), 2946-2957.

Legionella Becoming a Mutualist: Adaptive Processes Shaping the Genome of Symbiont in the Louse *Polyplax serrata*

Jana Říhová[1], Eva Nováková[1,2], Filip Husník[1], and Václav Hypša[1,2,*]

[1]Department of Parasitology, University of South Bohemia, České Budějovice, Czech Republic

[2]Biology Centre, Institute of Parasitology, CAS, v.v.i., České Budějovice, Czech Republic

*Corresponding author: E-mail: vacatko@prf.jcu.cz.

Accepted: October 20, 2017

Data deposition: This project has been deposited at GenBank under the accession CP021497.

Abstract

Legionellaceae are intracellular bacteria known as important human pathogens. In the environment, they are mainly found in biofilms associated with amoebas. In contrast to the gammaproteobacterial family Enterobacteriaceae, which established a broad spectrum of symbioses with many insect taxa, the only instance of legionella-like symbiont has been reported from lice of the genus *Polyplax*. Here, we sequenced the complete genome of this symbiont and compared its main characteristics to other *Legionella* species and insect symbionts. Based on rigorous multigene phylogenetic analyses, we confirm this bacterium as a member of the genus *Legionella* and propose the name *Candidatus* Legionella polyplacis, sp.n. We show that the genome of *Ca.* Legionella polyplacis underwent massive degeneration, including considerable size reduction (529.746 bp, 484 protein coding genes) and a severe decrease in GC content (23%). We identify several possible constraints underlying the evolution of this bacterium. On one hand, *Ca.* Legionella polyplacis and the louse symbionts *Riesia* and *Puchtella* experienced convergent evolution, perhaps due to adaptation to similar hosts. On the other hand, some metabolic differences are likely to reflect different phylogenetic positions of the symbionts and hence availability of particular metabolic function in the ancestor. This is exemplified by different arrangements of thiamine metabolism in *Ca.* Legionella polyplacis and *Riesia*. Finally, horizontal gene transfer is shown to play a significant role in the adaptive and diversification process. Particularly, we show that *Ca.* L. polyplacis horizontally acquired a complete biotin operon (bioADCHFB) that likely assisted this bacterium when becoming an obligate mutualist.

Key words: symbiosis, horizontal gene transfer, genome evolution.

Introduction

Legionellaceae are mainly known as important human bacterial pathogens (Diederen 2008), although their life strategy is generally bound to biofilms where they live in intracellular symbiotic associations with amoebas and other protists (Fields 1996). Consequently, all known species have therefore been described either from water sources or clinical materials. The only known exception to this rule is the obligate symbiont of the rodent lice *Polyplax serrata* and *P. spinulosa* originally characterized by light and electron microscopy (Ries 1931; Volf 1991). Based on the 16S rDNA sequence, this bacterium was later on suggested to be a member of the genus *Legionella* (Hypsa and Krizek 2007). In this work, the transition from a typical legionella to an obligate symbiont was inferred from the presence of the bacterium in all tested louse individuals, suggesting their transovarial transmission, and from the typical shift in GC content. These traits are common for many obligatory symbionts living intracellularly in various insects. Typically, such bacteria reside in specialized organs, usually called bacteriomes, and are presumed to supply the host with some essential compounds, mainly vitamins and amino acids, missing in their diet (Douglas 1989). Their evolution/adaptation toward this role is usually accompanied by extensive modifications of their genomes, consisting mainly of gene (or complete function) losses and decrease of GC content (Woolfit and Bromham 2003). Although the latter is considered a sign of relaxed selection, the reduction of metabolic capacities results mostly from adaptive processes. In the course of evolution, these bacteria lose metabolic functions fulfilled by the host and retain (sometimes even acquire by horizontal transfer; Husnik et al. 2013; Nakabachi et al. 2013; Nikoh et al. 2014) functions essential for the host's development and/or

reproduction. Due to these processes, the bacteria can evolve an array of different life strategies, from parasites/pathogens to obligate symbionts. It was for example demonstrated that the notorious insect parasite *Wolbachia* turned into a mutualistic bacterium in the bedbug *Cimex lectularius* upon horizontal acquisition of an operon for biotin biosynthesis (Nikoh et al. 2014; Gerth and Bleidorn 2017).

Several decades lasting research on insect-bacteria symbiosis has revealed a broad variety of these associations and has resulted in many fundamental discoveries (Moran 1996; Ochman and Moran 2001; McCutcheon and Keeling 2014; Moran and Bennett 2014). One of the interesting outcomes is the broad diversity of bacteria which are generally capable to establish obligate symbiosis with insects, including such diverse taxa as Enterobacteriales, Rickettsiales, Blattabacteriaceae, and Flavobacteriales. However, within this broad bacterial diversity, several groups show a particular tendency to this behavior (e.g., Enterobacteriaceae, more specifically the genera *Arsenophonus* and *Sodalis*), whereas others are found less frequently. Legionellaceae is one of these rare groups, with the *Polyplax*-associated bacterium being the only case of such intracellular obligate insect symbiont. This raises an interesting question on how this bacterium established its unique symbiotic relationships with lice and what changes it underwent during its adaption to obligate symbiosis. In this respect, it is important that there is a solid body of genomic information for both the genus *Legionella* and many other bacterial symbionts from various blood sucking insects. In a recent genomic work, comparing 38 *Legionella* species, Burstein et al. (2016) revealed a surprising diversity of their genomic traits, including major differences in the most general characteristics such as genome size or GC content. The main focus of this genomic comparison, particularly important in respect to the evolution of legionellae virulence, were effectors, especially those responsible for entering the host cell. The computational analysis inferred several thousands of candidate effectors and indicated that many of these factors have been recently acquired by horizontal transfer. This shows the genus *Legionella* as a highly dynamic system with ongoing genome rearrangements and adaptations toward the intracellular lifestyle. The newly characterized legionella-like louse symbiont thus provides us with a unique opportunity to study the processes of the genomic changes and compare them with "free living" *Legionella* on one hand and to the unrelated enterobacterial symbionts on the other hand.

In this study, we approach this issue by exploring genomic features of the symbiont from *P. serrata*, a louse associated with several species of the genus *Apodemus* (wood mice) across Europe (Stefka and Hypsa 2008). Among the 78 described species of the cosmopolitan genus *Polyplax*, associated exclusively with rodents and insectivores (Durden and Musser 1994), *P. serrata* and closely related species *P. spinulosa*, are the only two species for which the presence of symbiotic bacteria (and their origin within the genus *Legionella*) has so far been determined (Hypsa and Krizek 2007). Since all sucking lice rely on the limited resources of the mammalian blood, the presence of similar obligatory symbionts in other *Polyplax* species is very likely, but their taxonomic/phylogenetic positions are difficult to anticipate. Generally, different groups of lice were shown to host different taxa of symbiotic bacteria (Allen et al. 2016) which provides evidence for several replacements or independent acquisitions of the symbionts during the lice evolution. Using complete genomic data, we now confirm the unique phylogenetic origin of the *P. serrata* symbiont within the genus *Legionella*. We also show that its genome followed an evolutionary route typical for obligatory symbionts, and underwent a surprisingly convergent evolution with the unrelated enterobacterial symbiont from the hominid lice. Finally, we show that its adaptation to the role of obligatory symbiont included horizontal acquisition of six coding genes in a biotin operon, in analogy to the mutualistic *Wolbachia* from bedbugs.

Materials and Methods

Sample Preparation

The specimens of *Polyplax serrata* were collected during autumn 2011 from *Apodemus flavicollis* mice trapped around Baiersbronn, Germany and stored in absolute ethanol at $-4\,^{\circ}C$. Since the obligate bacterial symbionts are uncultivable outside the host cells, total DNA was obtained by extraction from whole abdomens of 25 louse individuals (QiaAmp DNA Micro Kit, Qiaqen). DNA concentration was assessed by the Qubit High Sensitivity Kit (Invitrogen) and 1% agarose gel electrophoresis.

Genome Sequencing and Assembly

The *Polyplax* sample was sequenced on one lane of Illumina HiSeq2000 (GeneCore, Heidelberg) using 2×100 paired-end reads (PE) library with an insert size of 150 bp. After quality checking and filtering in BBtools (https://jgi.doe.gov/data-and-tools/bbtools/; last accessed October 30, 2017), the resulting data set contained 309,892,186 reads in total. The reads were assembled using the SPAdes assembler v 3.10 (Bankevich et al. 2012), under default settings with the parameter *careful*, decreasing number of mismatches and indels. To check for possible presence of bacterial plasmid(s) in the data, we submitted complete assembly to the PlasmidFinder (Carattoli et al. 2014) with sensitivity set to three different thresholds (95%, 85%, and 60%). Phylogenetic affiliations of the contigs were determined by PhylaAMPHORA (Wu and Scott 2012). Of the total 124,985 contigs, 112 were assigned to the order Legionellales. Trimmed reads were mapped on these contigs and filtered using BWA v. 0.7.15 (Li and Durbin 2009) and retrieved by Samtools (Li et al. 2009). This set of reads was subsequently assembled by two alternative assemblers,

SPAdes and A5 assembly pipeline (Coil, Jospin and Darling 2015). The latter software produced a closed 529,746-bp long genome. Its quality was checked and base calls were polished by Pilon v1.20 (Walker et al. 2014).

Genome Annotation

The genome was annotated using two tools services: RAST (Aziz et al. 2008), and PROKKA (Seemann 2014). The complete genome was deposited in GenBank with the accession number CP021497. To scan for potential horizontal gene transfer(s) (HGT), we retrieved the 50 most similar sequences for each protein using the BLASTp algorithm (Altschul et al. 1990) against the nr (nonredundant) protein database. Metabolic pathways for B vitamins and cofactors were reconstructed using KEGG Mapper (http://www.genome.jp/kegg/mapper.html; last accessed October 30, 2017) and EcoCyc database (https://ecocyc.org/; last accessed October 30, 2017). The absence of genes in important metabolic pathways was verified using BLAST searches. The number of proteins with transmembrane helices was predicted in TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/; last accessed October 30, 2017) with the default parameters. The signal peptides were detected by SignalP online server (http://www.cbs.dtu.dk/services/SignalP/; last accessed October 30, 2017) with the cutoff values estimated by the program. The presence of CRIPSR repeats was checked using Geneious software (Katoh and Standley 2013). The origin and terminus of replication was determined in GenSkew online tool (http://genskew.csb.univie.ac.at/; last accessed October 30, 2017) according to the formula: GC skew = (G−C)/(G + C) and with the genome split into 9,000 windows.

Phylogenetic Analyses

Sixty-four proteins for a multigene analysis were selected based on the recent genomic study by (Burstein et al. 2016) and their orthologs in the *L. polyplacis* were determined using BLASTp search against the *L. polyplacis* genome. The sequences were aligned using MAFFT v. 1.3.5 (Katoh et al. 2002) implemented in the Geneious software. Equivocally aligned positions and divergent regions were eliminated by GBlocks (Castresana 2000) under the less stringent selection options. The alignments were created separately for each protein and multigene matrix was obtained by their concatenation. To minimize phylogenetic artifacts caused by rapid evolution and nucleotide bias of the symbiont sequences, we used PhyloBayes MPI v. 1.5a (Lartillot et al. 2013) with the CAT-GTR model and dayhoff6 amino acid recoding, and ran it for 32,000 generations. This approach has been previously shown to decrease phylogenetic artifacts affecting branching of symbiotic bacteria in bacterial phylogenies (Husnik et al. 2011).

For the candidate HGTs, that is, the biotin operon genes (see Results and Discussion), we prepared a representative set of orthologs covering several bacterial groups and used the alignment method described earlier to build amino acid matrices. The best evolutionary model for all matrices, determined in Prottest 3.2 (Darriba et al. 2011) by Akaike information criterion (AIC), was LG with a proportion of invariable sites and evolutionary rates separated in four categories of gamma distribution (LG + I+G). Phylogenetic reconstructions were done by maximum-likelihood analyses with 100 bootstrap replicates using PhyML v. 2.2.0 (Guindon et al. 2010) for each gene separately and also for a concatenated matrix of all six genes. Posterior probabilities for individual branches were determined by MrBayes v 3.2.6 (Huelsenbeck and Ronquist 2001) with the same evolutionary model (LG + I+G) and remaining parameters determined by the analysis. The analysis was run under the default four-chain setting for 10,000,000 generations. Convergence of all Bayesian analyses was checked in Tracer v1.6.0 (Rambaut et al. 2014), and for MrBayes also by the values of standard deviation of split (<0.01) and PSRF + (reached the value 1.0). Based on the availability for each analyzed gene, one of the following bacteria was used as an outgroup: *Kurthia* sp. (Firmicutes), *Geobacter sulfurreducens* (Deltaproteobacteria), or *Cyanothece* sp. (Cyanobacteria). Graphical representation of the trees was processed in FigTree (http://tree.bio.ed.ac.uk/software/figtree/; last accessed October 30, 2017) and Inkscape (http://www.inkscape.org/; last accessed October 30, 2017).

Comparative Genome Analyses

In order to reveal general similarity patterns and possible functional convergences between the symbiont of *P. serrata* and other symbiotic bacteria, we have treated the genomes and genes as communities and their components, respectively, and employed nonmetric multidimensional scaling (NMDS). This clustering analysis, routinely used in microbial ecology, produces an ordination based on a distance or dissimilarity matrix (Legendre and Legendre 2012). The dissimilarity matrix based on Bray–Curtis distances was calculated for 4,632 GOCs from 94 bacterial genomes (supplementary table S2 in supplementary file S1, Supplementary Material online). Draft genomes of poor quality have not been included. In particular we have omitted two genomes of high interest, that is, *Riesia pediculischaeffi* and *Sodalis* endosymbiont of *Proechinophthirus fluctus* (Boyd et al. 2016). All the genome data were retrieved as proteomes from GenBank and assigned to particular functional orthologs as defined by the COG database released December 2014 (Tatusov et al. 2000; Galperin et al. 2015) using rpsBLAST (Altschul et al. 1990) limited by a single hit and a maximal e-value set to 10^{-5}. Although the data set includes "free living" Legionellaceae with large genomes, the matrix was transformed using a percentage proportion for each COG ortholog calculated from the COG sum of each genome. The matrix of 4,632 orthologs was converted into the biom format. The Bray Curtis dissimilarities were calculated and the NMDS analysis was performed using vegan package

Table 1
Comparison of the Main Genomic Characteristics[a]

	Legionella spp.	*Legionella polyplacis*	Endosymbionts of Hominid Lice[b]
Genome size (bp)	2,367,087–4,818,052	529,746	528,700–580,415
Number of proteins	1,984–3,986	484	444–529
CG content (%)	36.60–51.10	23	23.75–31.80
Number of COGs	1578–2603	452	428–508
ø coding density (%)	85.43–93.10[c]	86.50	79.17–85.96
ø gene size (bp)	744–1,113	931	849–896

[a]The source data are provided in supplementary table S1 in supplementary file S1, Supplementary Material online.
[b]The *Riesia phthiripubis* genome was not included into the comparison due to the presence of 444 psedogenes in the NCBI annotation (CP012846.1).
[c]Coding densities were calculated only for the species with complete annotations available.

functions in R (Dixon 2003). In order to show the genome functional similarities among lice symbionts with distinct phylogenetic origin, a Neighbour Joining (NJ) tree was calculated in T-rex (Alix et al. 2012) using the same distance matrix.

Proposal for the Species Name of Legionella-like Endosymbiont

As we show that the symbiont clusters within the genus *Legionella*, we propose in accordance with the terms for species designation which have not been cultivated in a laboratory media the name "*Candidatus* Legionela polyplacis" sp. nov. (hereafter *Legionella polyplacis* for simplicity). The specific name "polyplacis" refers to the genus of its insect host, the louse *Polyplax serrata*. The bacterium is the only member of the genus for which endosymbiotic association with insects has been documented.

Results and Discussion

Basic Genomic Properties

The complete genome of *Legionella polyplacis* is 529,746 bp long, has an extremely low CG content (23%) and a coding density 84.8%. According to the RAST annotation (supplementary table S3 in supplementary file S1, Supplementary Material online), it contains 484 protein encoding genes (pegs) with the average length of 939 bp, 3 genes coding for rRNAs and 36 genes for tRNAs. In two cases, pairs of adjacent pegs, annotated by identical names, corresponded clearly to a gene interrupted by a stop codon (DNA polymerase I [EC 2.7.7.7] and COG1565: Uncharacterized conserved protein). Since we confirmed the presence of the stop codons by mapping the reads on the assembled genome, we suppose that these cases represent either a functional split into two separate genes or an early stage of pseudogenization. To avoid an arbitrary determination of other possible pseudogenes, based on the sequence lengths, we provide in supplementary table S3 in supplementary file S1, Supplementary Material online, the comparison of length ratios between the *L. polyplacis* genes and their closest BLAST hits. In summary, of the 484 coding genes, only 48 are shorter than 90% of the BLAST-identified closest homologs, 23 of them shorter than 80%. The genome size and gene number place *L. polyplacis* among many other obligate symbionts of insects. In table 1, we show the comparison of these general genomic characteristics between typical legionellae, *L. polyplacis* and other hominid louse symbionts. The positions of the origin and terminus of DNA replication are shown on the GC-skew plot (supplementary file S4, Supplementary Material online), corresponding to the minimum and maximum values, respectively. The origin position coincides with the location of *dnaA* gene.

Of the coding genes, 105 contain transmembrane helices and 4 signal peptides. No CRIPSR repeats were identified. PlasmidFinder, run under the sensitivity range 95–60%, did not reveal any candidate plasmid sequence. Majority of the coding genes (445 genes) could be assigned by BLASTp unequivocally to the genus *Legionella* (i.e., 5 best hits corresponded to the genus *Legionella*), for 20 genes the search returned *Legionella* together with some other bacterial genera, and for 30 genes we either did not get any match or the hits corresponded to other bacteria than *Legionella*. Among the "non-legionella" genes, some were highly conserved (ribosomal proteins) or very short genes difficult to assign based on the BLAST algorithm. However, six of these "non-legionella" genes formed a complete biotin operon known from several nonrelated symbiotic bacteria, suggesting its horizontal acquisition in the *L. polyplacis*). A significance of this HGT in respect to the symbiotic nature of this legionella is discussed below. The genes with no BLAST hits were mostly annotated as hypothetical proteins, usually with very short sequences. Of the metabolic functions considered particularly significant in the endosymbiotic bacteria, that is, biosynthesis of the vitamins and cofactors, we detected several complete pathways (fig. 1) and the horizontally acquired capability of biotin synthesis (fig. 2).

Origin and Phylogeny

The multigene analyses confirm that *L. polyplacis*, the symbiotic bacterium from the lice of the genus *Polyplax*, have

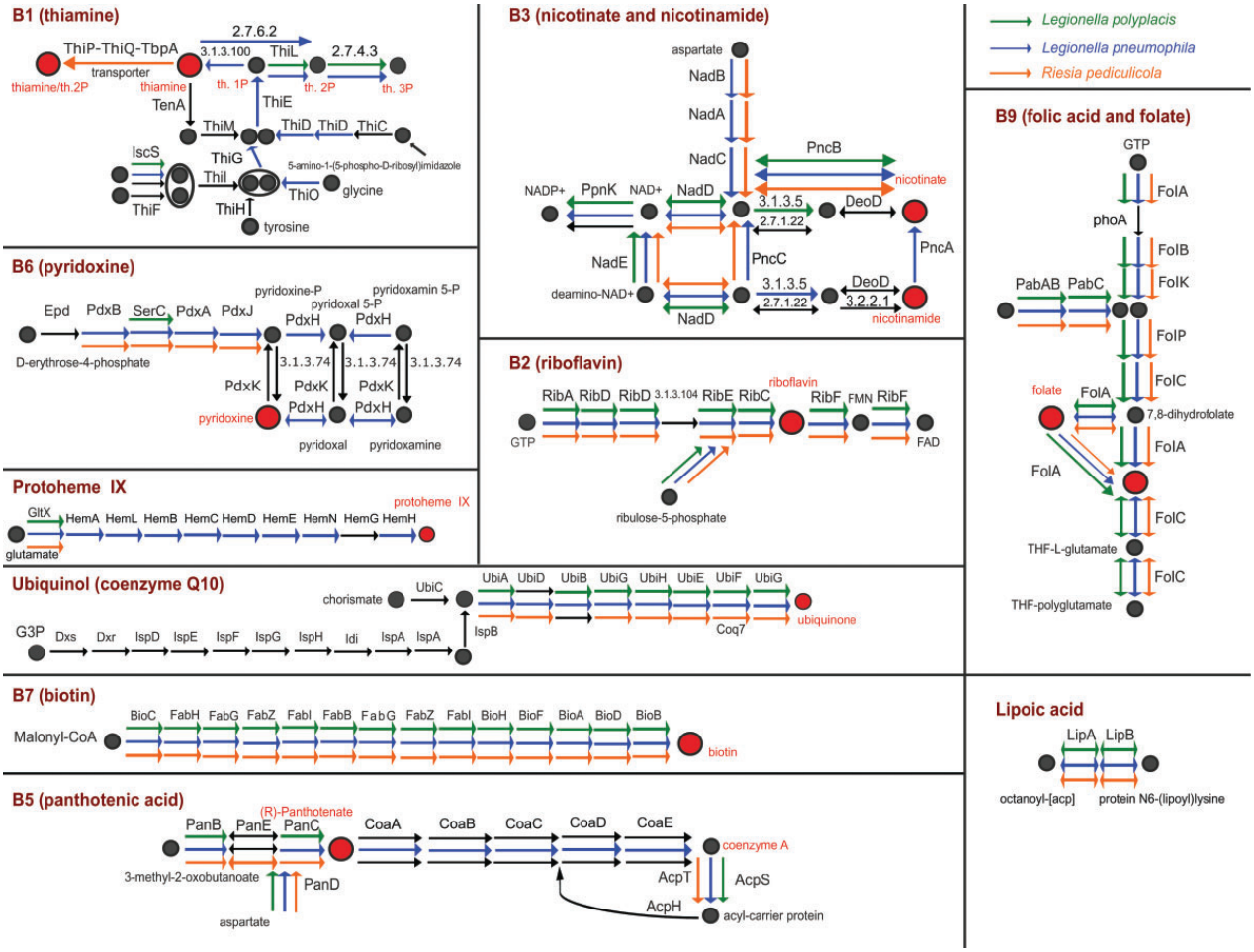


Fig. 1.—Comparison of B-vitamins and cofactors pathways for *Ca*. Legionella polyplacis, *Riesia pediculicola*, and *Legionella pneumophila*. Black arrows designate missing genes. Ellipses around two metabolites highlight the same following steps in the pathway.

indeed originated within the genus *Legionella* (fig. 3 and supplementary file S2, Supplementary Material online), as previously suggested based on 16S rDNA (Hypša and Křížek 2007). Although in the PhyloBayes analysis the chain did not converge even after 32,000 generation, the maximum likelihood PhyML analysis yielded an identical result in respect to the *L. polyplacis* position. In both analyses, it clustered on an extremely long branch, but with a strong support, within the group corresponding to the *L. micdadei* clade of Burstein et al. (2016). This position was further supported by a BLAST analysis of the genes conserved in all species, including the symbiont, where the first eight hits belonged to the *L. micdadei* clade (i.e., *L. feeleii, L. lansingensis, L. brunensis, L. hackeliae, L. jamestowniensis, L. maceachernii, L. jordanis*, and *L. nautarum*). The basic structure of the whole *Legionella* cluster (i.e., monophyly of the main groups) corresponds to that reported by Burstein et al. (2016). The PhyML-derived topology retains closer similarity to their ML-based tree. This general agreement among results of our ML, Bayesian inference based on the recoded matrix, and the Burstein's et al. (2016) topologies indicate that the data provide a reliable information for reconstructing the relationships within this group.

The position of *L. polyplacis* deep within the *Legionella* phylogeny clearly indicates that it most likely evolved from a "free living" ancestor with similar ecology to other legionellae. Considering the capability of the legionellae species to switch between symbiotic forms living in amoebas and pathogenic forms infecting mammal cells, it is difficult to hypothesize which of these forms gave rise to the mutualistic *L. polyplacis*. Although the *L. micdadei* clade possess several distinctive features when compared with the *L. pneumophila* clade, for example, smaller genomes, lower number of effectors, general tendency to loosing genes (Burstein et al. 2016), their possible evolutionary significance is unclear. More generally, due to the considerable reduction of *L. polyplacis* genome and our current tools for exploring this association, it is impossible to deduce which of the inherited genetic factors and traits could play role in transition toward the nutritional

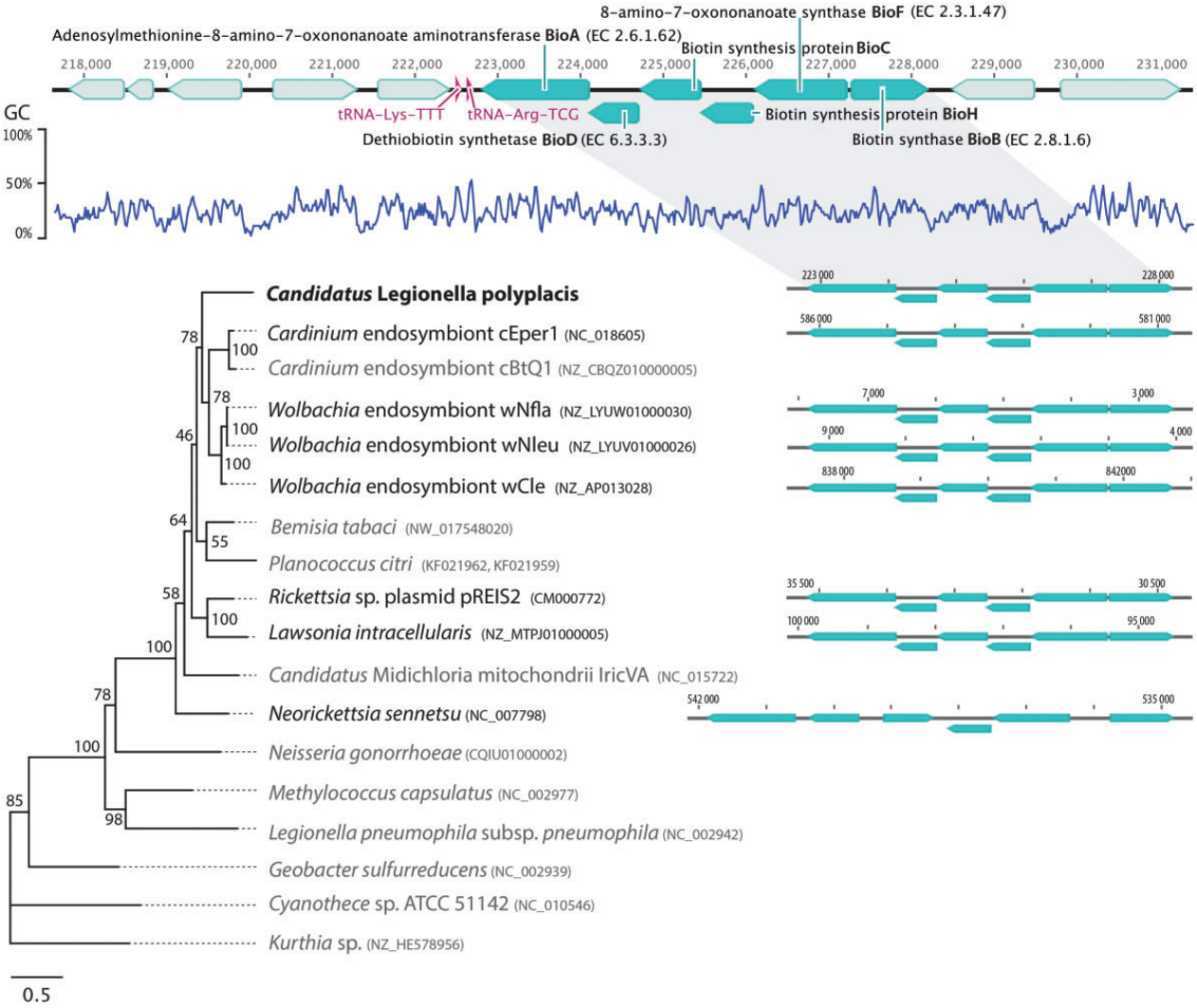


Fig. 2.—Structure of the horizontally acquired biotin operon in *Ca*. Legionella polyplacis and its putative evolutionary origin, based on ML analyses of all available genes. Blue arrowhead blocks represent genes for biotin synthesis organized in an intact operon. The adjacent genes code for the following proteins: for BioA from right to left: hypothetical protein, Isopentenyl-diphosphate delta-isomerase FMN-dependent, 4-hydroxy-tetrahydrodipicolinate synthase, 4Fe-4S ferredoxin, TsaB protein; for BioB from left: Aminomethyltransferase (glycine cleavage system T protein), and Glycine dehydrogenase [decarboxylating] (glycine cleavage system P2 protein). The position of the complete operon is visualized in the respective genomes (black font). The species with missing genes or disrupted operon structure (*Legionella pneumophila*) are in gray. The numbers at the tree nodes stand for the bootstrap values.

insect symbionts and how much were these factors taxonomically specific. For the same reason, no meaningful differential comparison could be done for *L. polyplacis* and other *Legionella* species.

Genomic Evolution

Comparing the genome size of *L. polyplacis* (529,746 bp and 484 coding genes) to that determined for other legionellae shows that similarly to other insect symbionts, the genome of *L. polyplacis* experienced considerable reduction in course of its adaptation to the symbiotic lifestyle. Although the 38 *Legionella* spp. compared by Burstein et al., (2016) varied considerably in their genome sizes (2–5 Mbp; 2,000–4,500 genes), they do not approach the degree of reduction determined for the *L. polyplacis* genome. Due to this dramatic reduction, *L. polyplacis* lost many complete pathways and systems present in other *Legionella* spp. Since provisioning B-vitamins and cofactors is considered as fundamental function of symbionts in blood-sucking insects (Nogge 1981; Hosokawa et al. 2010; Snyder et al. 2010), we assessed this metabolic capacity of *L. polyplacis* in comparison to its phylogenetic relative *Legionella pneumophila* and phylogenetically distant louse symbiont *Riesia pediculicola* (fig. 1). All three

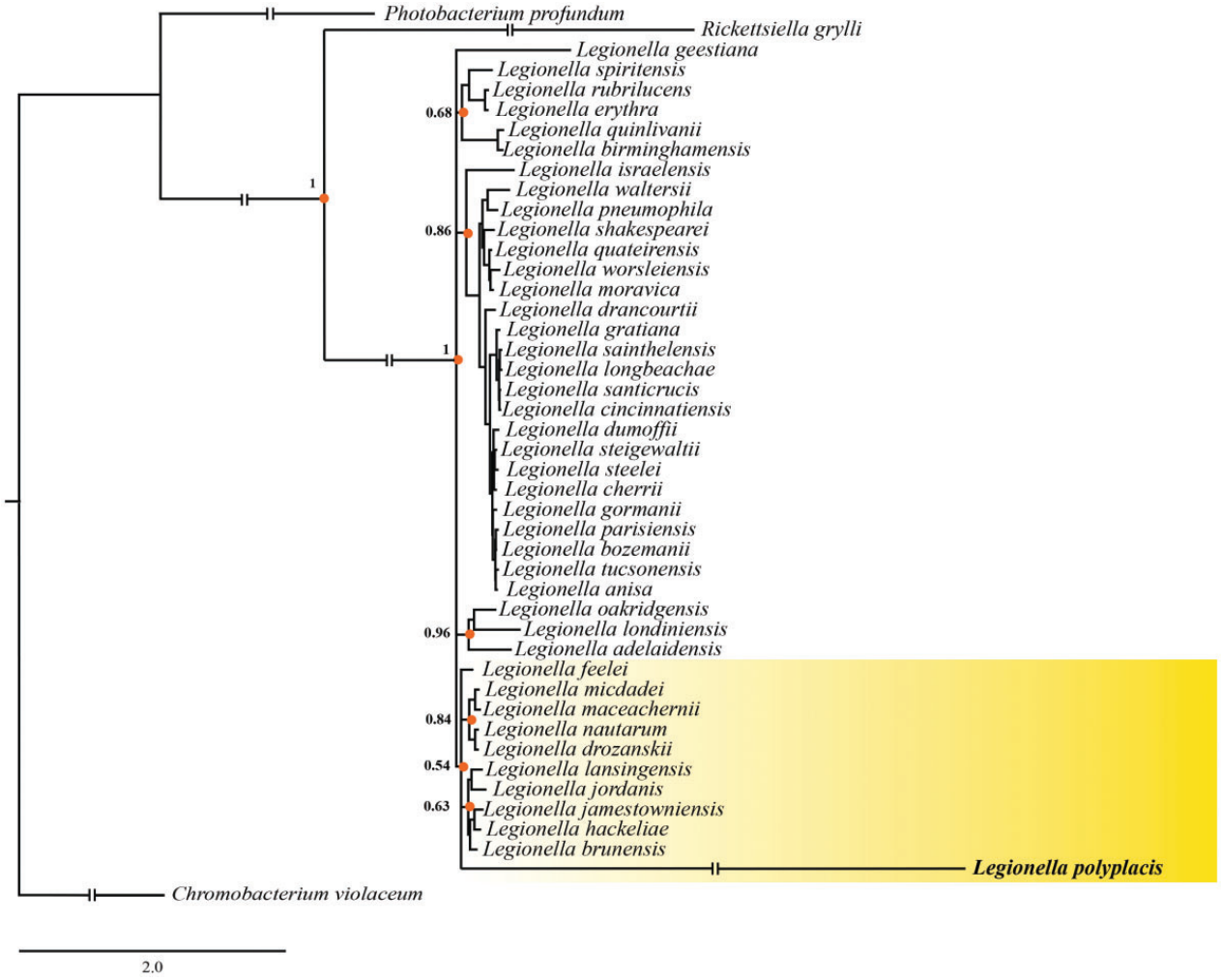


Fig. 3.—Phylogenetic tree inferred by PhyloBayes analysis of concatenated 64-gene matrix. The orange dots (with posterior probability values) highlight arbitrary selected monophyletic clusters as main components of the tree structure. The interrupted branches were shortened by 50%.

bacteria possess complete pathways for riboflavin (B2), biotin (B7), lipoic acid, and folate (B9); the last one without a known enzyme responsible for removing the triphosphate motif from 7,8-dihydroneopterin 3′-triphosphate. The rest of the pathways are in various stages of incompleteness/degradation, as demonstrated in figure 1. Perhaps the most striking example of genomic degradation is a complete lack of virulence-related secretion systems in the *L. polyplacis* genome. The "free living"/pathogenic legionellae, like other intracellular bacteria, possess molecular machinery for entering the host cell, surviving, and reproducing. Major components of this machinery are the Dot/Icm type IVB secretion system, Lsp type II secretion system, and depending on the species, also various forms of the type IVA secretion systems (Joseph et al. 2016). None of these systems is encoded by the *L. polyplacis* genome, and its only remaining secretion machinery is thus the Sec-SRP system. This is in line with the suggested characteristic of *L. polyplacis* as an obligate mutualist. This type of symbionts are highly adapted to their host and in contrast to the facultative secondary symbionts (Masui et al. 2000; Dale et al. 2001; Wilkes et al. 2010), they often do not use secretion systems (Perez-Brocal 2006). Another capability often lost or compromised in the insect bacterial symbionts is the cell wall formation and shape determination (Moran et al. 2008). In supplementary table S4 in supplementary file S1, Supplementary Material online, we show a more detailed comparison of the *L. polyplacis* genome capacity to that of *L. pneumophila*, *R. pediculicola*, and *Wigglesworthia glossinidia*. To make the overview also comparable to other bacterial symbionts, we followed the pathways/genes list published by Moran et al. (2008) for a broader taxonomic range of insect symbionts. As shown in this table, *L. polyplacis* retains the capacity to synthesize peptidoglycan, but it lost the genes required for building the lipopolysacharide outer layer,

particularly for the lipid A biosynthesis. However, compared with the "free living" legionellae, the peptidoglycan pathway in *L. polyplacis* lacks penicillin binding protein class A (aPBP). Interestingly, the same applies to the *R. pediculicola* genome. In addition, both *L. polyplacis* and *R. pediculicola* possess all four genes required for rod shape determination (ftsZ, ispA, mreB, rodA). Considering the high degree of completeness of the cell wall formation, it is likely that the function of missing aPBP is substituted by another enzyme. Such possibility is in line with the most recent discoveries showing that some of the cell wall formation functions can be fulfilled by the SEDS (shape, elongation, division, and sporulation) proteins (Meeske et al. 2016). More specifically, in *Bacilus subtilis* the function of the aPBP was shown to be replaced by the rodA enzyme, also present in the *L. polyplacis* and *R. pediculicola* genomes. Among the other major systems, a significant degradation occurred in the biosynthesis of amino acids with no complete pathway retained, and in the recombination/repair system with a high proportion of lost genes (supplementary table S4 in supplementary file S1, Supplementary Material online).

Both of the characteristics discussed earlier, that is, the strong compositional bias of the sequences toward GC, and considerable genome reduction, including loss of complete pathways and metabolic competencies, clearly place *L. polyplacis* among other obligatory mutualists of insects, despite the absence of a direct functional evidence.

Since nothing is known about the origin of bacterial symbionts in *Polyplax* species other than *P. serrata* and *P. spinulosa*, it is difficult to estimate the time frame of this reduction. In phylogenetic study published by Light et al., (2010), the family Polyplacidae are revealed as a paraphyletic taxon, with the genus *Polyplax* branching as sister group to a Pediculidae/Phthiridae/Pedicinidae cluster. The estimated diversification time for these two groups, ~45 Ma, can thus be considered an upper limit for the origin of *L. polyplacis*. This would place *L. polyplacis* close in age to another obligatory symbiont associated with blood sucking insect, namely *Wigglesworthia glossinidia*, with the estimated origin of 40 Ma (Naito and Pawlowska 2016). However, since the *Polyplax*'s long branch in the Light et al. (2010) analysis indicates that other closer relatives of *Polyplax* might be missing in the taxa set, this time frame may be considerably overestimated. For example, another louse symbiont, *Riesia pediculicola* from the human louse, was shown to had reached similar genome state within ~13–25 Ma (Boyd et al. 2014).

Regardless the uncertainty in time estimates, *L. polyplacis*, *R. pediculicola*, and the other recently described symbionts of lice (Boyd et al. 2017) reached similar basic genome characteristics (table 1). More interestingly, the NMDS analysis, depicturing the overall similarities among the genomes (fig. 4), and particularly the distance based NJ tree (fig. 5) placed *L. polyplacis* far from other legionellae but remarkably close to the phylogenetically unrelated lice symbionts from the genus *Riesia* (i.e., member of the *Arsenophonus* cluster). In the NJ tree, all of the louse symbionts, including *Puchtella*, even form a monophyletic lineage. It is important to notice that the clustering is not determined by the genome size as the other highly modified symbionts are scattered across the whole NMDS plot. Even more interestingly, the primary symbionts associated with blood feeding pupiparans, that is, *Wigglesworthia* spp., *Arsenophonus melophagi*, and *A. lipopteni*, cluster within a distant independent group relatively close to each other. In other words, the symbionts of blood feeding insects show clear tendency toward clustering according to their host phylogeny (and hence perhaps biology) rather than their own phylogeny. Although the sample of the symbionts from the blood-feeding insects is small for any decisive inference, the pattern suggests that the clustering reflects the convergence in the genome contents and consequently their functions rather than a general genome reduction and the host's source of nutrients.

Metabolism, Adaptive Processes, and HGT

To some degree, however, the convergent evolution/adaptation described earlier, is limited by phylogenetic constraints, that is, availability of different metabolic machineries inherited from unrelated bacterial ancestors. For example, both *L. polyplacis* and the *Riesia* species lost capacity to synthetize thiamin. In *Riesia*, this incapacity is compensated by a specific thiamin ABC transporter inherited from its ancestor (as inferred from the presence of ABC thiamin transporter in other *Arsenophonus* spp.). Boyd et al. (2017) hypothesized that the loss of thiamin synthesis by *Riesia* in hominid lice and its retention by *Puchtella* in colobus monkey lice may reflect diet differences of the two mammal hosts. They suggest that the complex diet of the hominids makes thiamin available for scavenging by *Riesia*. According to this view, *L. polyplacis* from the rodent associated lice could also scavenge thiamin from its host. However, the specific thiamin transporter is not present in the known *Legionella* genomes and *L. polyplacis* thus lacks both the synthetic pathway as well as specific transporter. Using the CoFactor database (Fischer et al. 2010), we determined thiamine as the essential compound required as a cofacor by at least two of the *L. polyplacis* enzymes (transketolase and pyruvate dehydrogenase). It is therefore likely that *L. polyplacis* utilizes some alternative system for its acquisition. As it is known that the ABC transporters can bound to more ligands (ter Beek et al. 2014), we suggest that among the possible candidates for this role are the other transporters which could be adapted for nonspecific transfer. For example, the ability to bind thiamin is known for the ABC putrescine transporter (ter Beek et al. 2014). This transporter is present in *L. polyplacis* and was obviously inherited from its legionellae ancestor. Apart from the ABC putrescine transporter subunits, only 11 additional sequences were identified as potential transporters using the TCDB (Saier et al. 2006). Four of these genes were associated with

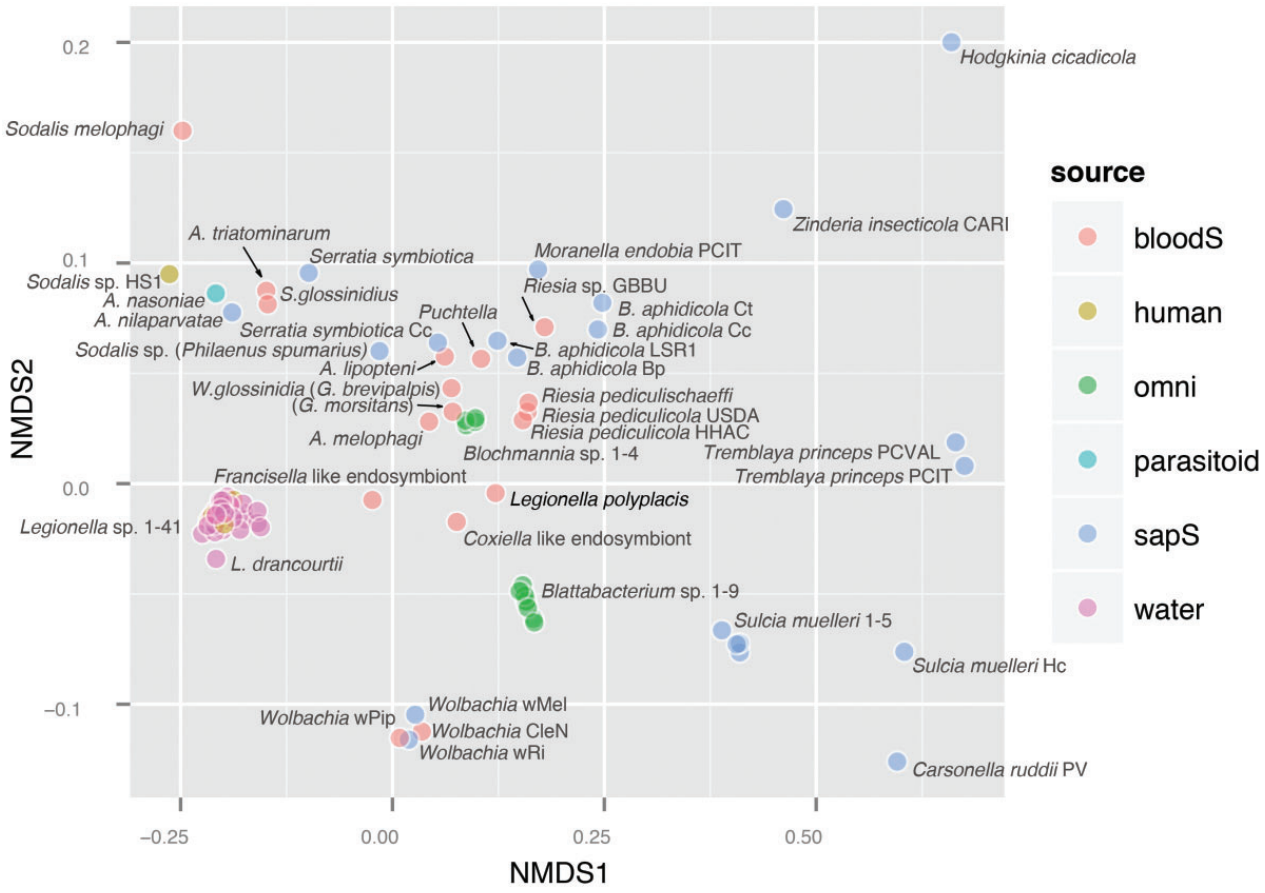


Fig. 4.—NMDS analysis based on Bray–Curtis dissimilarities calculated for the genome content of selected bacteria analyzed across all COG orthologs. The position of each genome represents as closely as possible the pairwise dissimilarity between the genomes. The legend abbreviations for the bacterial source are as follows: bloodS: blood sucking host, human: clinical isolate, omni: omnivorous insect host, parasitoid: that is, *Nasonia vitripennis* host, sapS: sap sucking insect host, water: environmental water sample. List of the analyzed genomes is provided in supplementary table S2 in supplementary file S1, Supplementary Material online.

cytochrome c biogenesis, two with sec-SRP system, and the additional two were identified as NhaA antiporter and phospho-*N*-acetylmuramoyl-pentapeptide transferase. Of the remaining three sequences, one was directly annotated as Major facilitator family transporter, whereas the affiliation of the two unannotated proteins (designated in supplemen tary table S3 in supplementary file S1, Supplementary Material online, as FIG00758517: hypothetical protein and putative transport protein) to the same transporter family was confirmed by the BLAST search.

A further important difference between *L. polyplacis* and *R. pediculicola* introduces horizontal gene transfer (HGT) as yet another determinant of a symbiont's evolutionary pathway. The significance of HGT for adaptation to a particular life style in bacteria is well recognized. For *Legionella*, Burstein et al. (2016) suggested that many of effectors, with possibly important functions in virulence and intracellular lifestyle, have been recently acquired by this mechanism. Within the system of thousands of predicted effectors, they were able identify only seven effectors shared by all tested *Legionella* species. This, together with the variability of main genome characteristics, shows *Legionella* as highly dynamic system capable of rapid adaptations. Unlike the "free living" legionellae, the HGT in *L. polyplacis* is bound to its symbiotic function known from other blood feeding insects, that is, provisioning the host with vitamins. In the case of B7 vitamin, this role in *L. polyplacis* is fulfilled by a horizontally acquired complete biotin operon with gene order conserved across the known homologues (fig. 2 and supplementary file S3, Supplementary Material online). Since only few instances of the complete homologous operon transfer are known from other bacteria, it would be difficult to establish a meaningful evolutionary scenario for this HGT and possible source of these genes in *L. polyplacis*. It is however interesting to note that the two most closely related operon homologs are also associated with insect symbionts, namely *Wolbachia* (strains wCle from

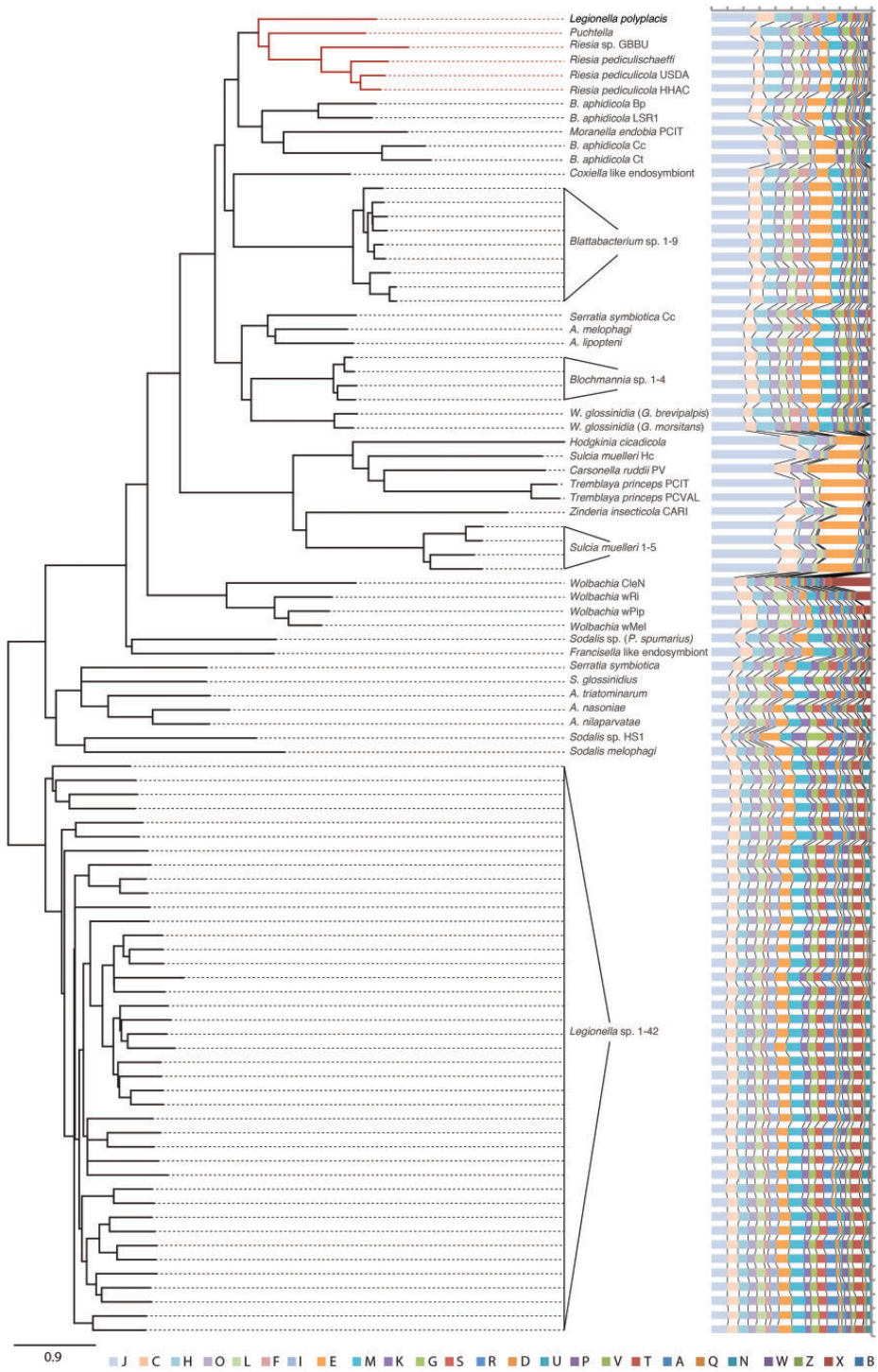


Fig. 5.—NJ tree calculated for Bray–Curtis distance matrix based on COG orthologs. The bar chart represents COG category content for individual genomes. The categories abbreviations are as follows: A: RNA processing and modification, B: Chromatin Structure and dynamics, C: Energy production and conversion, D: Cell cycle control and mitosis, E: Amino Acid metabolism and transport, F: Nucleotide metabolism and transport, G: Carbohydrate metabolism and transport, H: Coenzyme metabolism, I: Lipid metabolism, J: Translation, K: Transcription, L: Replication and repair, M: Cell wall/membrane/envelop biogenesis, N: Cell motility, O: Posttranslational modification, protein turnover, chaperone functions, P: Inorganic ion transport and metabolism, Q: Secondary Structure, T: Signal Transduction, U: Intracellular trafficking and secretion, V: Defense mechanism, W: Extracellular structures, X: Mobilome: prophages, transposons, Y: Nuclear structure, Z: Cytoskeleton, R: General Functional Prediction only, S: Function Unknown.

bed bugs and wNfla, Wnleu from bees) and *Cardinium* (strain cEper1 from parasitic wasps and cBtQ1 from whiteflies). In addition to the biotin operon, for which a rickettsial origin was suggested by Penz et al. (2012), *Cardinium* was shown to harbor high number of other HGT-acquired genes (Penz et al. 2012). Regarding possible source(s) of these HGTs, the authors suggested two main possibilities, both highly relevant in respect to this study. First, for part of the genes the closest homologs are associated with other insect symbionts. The second possible source are bacteria associated with amoebas, including *Legionella*. In addition, a considerable portion of the putative bacterial donors are capable to infect both insects and amoebas (Penz et al. 2012). Considering its presumable ancestral life style, that is, association with amoebas, and current symbiotic relationships with insects, the acquisition of biotin operon by *L. polyplacis* from either of these sources seems plausible hypothesis. In *Wolbachia* from the bed bug *Cimex lectularius*, the acquisition of the operon by HGT has been explicitly claimed to had enabled ecological transition of otherwise parasitic *Wolbachia* to an obligate mutualist (Nikoh et al. 2014). However, in this case, the parasitic ancestor of the symbiotic *Wolbachia* was likely to lack the biotin synthesis capacity altogether. Since legionellae generally possess the biotin synthesis capacity, the acquisition of the biotin operon by *L. polyplacis* appears as a replacement rather than a de novo acquisition of this function. However, in contrast to this compact six-gene biotin operon, in the genomes of "free living" legionellae BioC is separated from the rest of the genes. As a remnant of this arrangement, *L. polyplacis* retains the original BioC (position 269231-269893) apart from the complete horizontally transferred biotin operon (222799-228219). Regarding the absence of any known intermediate form between the pathogenic legionellae and *L. polyplacis*, it is difficult to hypothesize on the absence/presence of the biotin synthesis capacity in the *L. polyplacis* ancestor, and therefore on the possible role of this HGT in triggering its transition to symbiosis. Various circumstances could possibly drive *L. polyplacis* toward adaptive acceptance of this operon. For example, within the highly economized *L. polyplacis* genome, an arrangement of biotin synthesis within a single operon could prove more efficient than the genes scattered around the genome as in other legionellae. In addition, the acquisition of this operon could follow preceding loss of the biotin synthesis in *L. polyplacis* parasitic ancestor (similar to the transition from parasitic to mutualistic *Wolbachia* described earlier). Altogether, the patterns discussed earlier show how the combination of different evolutionary forces (phylogenetic constraint, adaptive pressure, and HGT) resulted in emergence of a unique genome/phenotype in the genus *Legionella*.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank Jana Martinu for providing the lice specimens. This work was supported by the Grant Agency of the Czech Republic (grant 14-07004S to V.H.). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

Literature Cited

Alix B, Boubacar D, Vladimir M. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Res. 40:W573–W579.

Allen J, Burleigh J, Light J, Reed D. 2016. Effects of 16S rDNA sampling on estimates of the number of endosymbiont lineages in sucking lice. Peerj 4:e2187.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Aziz R, et al. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Boyd B, et al. 2016. Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). Appl Environ Microbiol. 82(11):3185–3197.

Boyd BM, Allen JM, de Crecy-Lagard V, Reed DL. 2014. Genome sequence of *Candidatus* Riesia pediculischaeffi, endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee Lice. G3 Genes Genomes Genet. 4(11):2189–2195.

Boyd BM, et al. 2017. Primates, lice and bacteria: speciation and genome evolution in the symbionts of hominid lice. Mol Biol Evol. 34(7):1743–1757.

Burstein D, et al. 2016. Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. Nat Genet. 48(2):167–175.

Carattoli A, et al. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 58(7):3895–3903.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552.

Coil D, Jospin G, Darling A. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31(4):587–589.

Dale C, Young SA, Haydon DT, Welburn SC. 2001. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci U S A. 98(4):1883–1888.

Darriba D, Taboada G, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27(8):1164–1165.

Diederen B. 2008. *Legionella* spp. and Legionnaires' disease. J Infect. 56(1):1–12.

Douglas A. 1989. Mycetocyte symbiosis in insects. Biol Rev Camb Philos Soc. 64(4):409–434.

Durden L, Musser G. 1994. The sucking lice (Insecta, Anoplura) of the world – a taxonomic checklist with records of mammalian hosts and geographical distribution. Bull Am Museum Nat History 218:1–90.

Fields B. 1996. The molecular ecology of legionellae. Trends Microbiol. 4(7):286–290.
Fischer J, Holliday G, Thornton J. 2010. The CoFactor database: organic cofactors in enzyme catalysis. Bioinformatics 26(19): 2496–2497.
Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(D1):D261–D269.
Gerth M, Bleidorn C. 2017. Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. Nat Microbiol. 2:16241.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.
Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. 2010. *Wolbachia* as a bacteriocyte associated nutritional mutualist. Proc Natl Acad Sci U S A. 107(2):769–774.
Huelsenbeck J, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755.
Husnik F, Chrudimsky T, Hypsa V. 2011. Multiple origins of endosymbiosis within the Enterobacteriaceae (gamma-Proteobacteria): convergence of complex phylogenetic approaches. BMC Biol. 9:87.
Husnik F, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153(7):1567–1578.
Hypsa V, Krizek J. 2007. Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). Microb Ecol. 54(2):242–251.
Joseph S, et al. 2016. Dynamics of genome change among *Legionella* species. Sci Rep. 6:33442.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.
Katoh K, Standley D. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 62(4):611–615.
Legendre P, Legendre LFJ. 2012. Numerical ecology. Amsterdam: Elsevier. Third English edition. ISBN 978-0-444-53868-0.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, Genome Project Data Processing S, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.,
Light J, Smith V, Allen J, Durden L, Reed D. 2010. Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). BMC Evol Biol. 10:292.
Masui S, Sasaki T, Ishikawa H. 2000. Genes for the type IV secretion system in an intracellular symbiont, *Wolbachia*, a causative agent of various sexual alterations in arthropods. J Bacteriol. 182(22):6529–6531.
McCutcheon JP, Keeling PJ. 2014. Endosymbiosis: protein targeting further erodes the organelle/symbiont distinction. Curr Biol. 24(14):R654–R655.
Meeske A, et al. 2016. SEDS proteins are a widespread family of bacterial cell wall polymerases. Nature 537(7622):634.
Moran N, McCutcheon J, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 42:165–190.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A. 93(7):2873–2878.
Moran NA, Bennett GM. 2014. The tiniest tiny genomes. Annu Rev Microbiol. 68:195–215.
Naito M, Pawlowska TE. 2016. The role of mobile genetic elements in evolutionary longevity of heritable endobacteria. Mobile Genet Elem. 6(1):7791–7796.
Nakabachi A, et al. 2013. Defensive bacteriome symbiont with a drastically reduced genome. Curr Biol. 23(15):1478–1484.
Nikoh N, et al. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc Natl Acad Sci U S A. 111(28):10257–10262.
Nogge G. 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology 82:101–104.
Ochman H, Moran NA. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292(5519):1096–1099.
Penz T, et al. 2012. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. PLoS Genet. 8(10):e1003012.
Perez-Brocal V, et al. 2006. A small microbial genome: the end of a long symbiotic relationship? Science 314(5797):312–313.
Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available from: http://beast.community/tracer, last accessed October 30, 2017.
Ries E. (co-authors). 1931. Die symbiose der läuse und federlinge. Z Morphol Ökologie Tiere 20(2–3):233–367.
Saier MH Jr, Tran CV, Barabote RD. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res. 34(Database issue):D181–D186.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069.
Snyder AK, Deberry JW, Runyen-Janecky L, Rio RV. 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. Proc R Soc. 277(1692):2389–2397.
Stefka J, Hypsa V. 2008. Host specificity and genealogy of the louse *Polyplax serrata* on field mice, *Apodemus* species: a case of parasite duplication or colonisation? Int J Parasitol. 38(6):731–741.
Tatusov R, Galperin M, Natale D, Koonin E. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28(1):33–36.
ter Beek J, Guskov A, Slotboom D. 2014. Structural diversity of ABC transporters. J Gen Physiol. 143(4):419–435.
Volf P. 1991. Postembryonal development of mycetocytes and symbionts of the spiny rat louse *Polyplax spinulosa*. J Invert Pathol. 58(1):143–146.
Walker B, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.
Wilkes TE, et al. 2010. The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. Insect Mol Biol. 19(s1):59–73.
Woolfit M, Bromham L. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol. 20(9):1545–1555.
Wu M, Scott A. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28(7):1033–1034.

Associate editor: Richard Cordaux

Paper 11:

Novakova, E., Woodhams, D. C., Rodríguez-Ruano, S. M., Brucker, R. M., Leff, J. W., Maharaj, A., ... & Scott, J. (2017). Mosquito microbiome dynamics, a background for prevalence and seasonality of West Nile virus. *Frontiers in microbiology*, *8*, 526.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 April 2017
doi: 10.3389/fmicb.2017.00526

Mosquito Microbiome Dynamics, a Background for Prevalence and Seasonality of West Nile Virus

Eva Novakova[1,2†], Douglas C. Woodhams[3†], Sonia M. Rodríguez-Ruano[1], Robert M. Brucker[4], Jonathan W. Leff[5,6], Amin Maharaj[7], Amnon Amir[8], Rob Knight[8,9] and James Scott[7,10]*

[1] Faculty of Science, University of South Bohemia, Ceske Budejovice, Czechia, [2] Biology Centre of ASCR, Institute of Parasitology, Ceske Budejovice, Czechia, [3] Department of Biology, University of Massachusetts Boston, Boston, MA, USA, [4] Rowland Institute, Harvard University, Cambridge, MA, USA, [5] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [6] Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA, [7] Sporometrics Inc, Toronto, ON, Canada, [8] Department of Computer Science and Engineering, Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, USA, [9] Department of Pediatrics, University of California San Diego, La Jolla, CA, USA, [10] Division of Occupational and Environmental Health, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:
Suhelen Egan, University of New South Wales, Australia

Reviewed by:
Yuval Gottlieb, Hebrew University of Jerusalem, Israel
Elisabeth Margaretha Bik, uBiome, USA

***Correspondence:**
Douglas C. Woodhams
dwoodhams@gmail.com

[†] These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *30 August 2016*
Accepted: *13 March 2017*
Published: *04 April 2017*

Citation:
Novakova E, Woodhams DC, Rodríguez-Ruano SM, Brucker RM, Leff JW, Maharaj A, Amir A, Knight R and Scott J (2017) Mosquito Microbiome Dynamics, a Background for Prevalence and Seasonality of West Nile Virus. Front. Microbiol. 8:526. doi: 10.3389/fmicb.2017.00526

Symbiotic microbial communities augment host phenotype, including defense against pathogen carriage and infection. We sampled the microbial communities in 11 adult mosquito host species from six regions in southern Ontario, Canada over 3 years. Of the factors examined, we found that mosquito species was the largest driver of the microbiota, with remarkable phylosymbiosis between host and microbiota. Seasonal shifts of the microbiome were consistently repeated over the 3-year period, while region had little impact. Both host species and seasonal shifts in microbiota were associated with patterns of West Nile virus (WNV) in these mosquitoes. The highest prevalence of WNV, with a seasonal spike each year in August, was in the *Culex pipiens/restuans* complex, and high WNV prevalence followed a decrease in relative abundance of *Wolbachia* in this species. Indeed, mean temperature, but not precipitation, was significantly correlated with *Wolbachia* abundance. This suggests that at higher temperatures *Wolbachia* abundance is reduced leading to greater susceptibility to WNV in the subsequent generation of *C. pipiens/restuans* hosts. Different mosquito genera harbored significantly different bacterial communities, and presence or abundance of *Wolbachia* was primarily associated with these differences. We identified several operational taxonomic units (OTUs) of *Wolbachia* that drive overall microbial community differentiation among mosquito taxa, locations and timepoints. Distinct *Wolbachia* OTUs were consistently found to dominate microbiomes of *Cx. pipiens/restuans*, and of *Coquilletidia perturbans*. Seasonal fluctuations of several other microbial taxa included *Bacillus cereus, Enterococcus, Methylobacterium, Asaia, Pantoea, Acinetobacter johnsonii, Pseudomonas*, and *Mycoplasma*. This suggests that microbiota may explain some of the variation in vector competence previously attributed to local environmental processes, especially because *Wolbachia* is known to affect carriage of viral pathogens.

Keywords: *Aedes vexans*, *Wolbachia*, *Culex pipiens*, arbovirus, flaviviridae, disease ecology

INTRODUCTION

Metazoa harbor diverse microbial communities (microbiota) largely dominated by bacteria (Bordenstein and Theis, 2015; Yadav et al., 2015). The microbiota modifies the ability of a host to be affected by, and to transmit, pathogens. Thus, understanding the relationship between microbiota and arthropod disease vectors, including mosquitoes, may impact mitigation of emerging infectious diseases (Dennison et al., 2014; Van Treuren et al., 2015).

Recently emerging vector-borne diseases have been linked to the introduction of non-native insect vectors and to changing ecological conditions including climate, urbanization, and greater human intrusion into areas where vectors and pathogens prevail (Bonizzoni et al., 2013). However, it is not known whether vector competence (i.e., the ability to transmit pathogens) is shaped mainly by environmental conditions, genetic background of the insect vector, or by the vector microbiota. Environmental factors and vector genotype both affect insect body size (Alto et al., 2008) and immunity status (Murdock et al., 2013), two traits that affect pathogen transmission. The microbiota may also influence disease dynamics.

Recent studies indicate that each mosquito species harbors specific microbiota even when larvae are raised under common conditions (Coon et al., 2014; Brooks et al., 2016). This distinction holds even when host species share habitat and are closely related and morphologically indistinct (Muturi et al., 2016). However, environmental conditions can also influence the microbiota of insect disease vectors (e.g., Jones et al., 2010; Tchioffo et al., 2016). It remains to be clarified whether mosquito genotype, region, or season is dominant in structuring microbial communities. For example, do differences in bacterial communities among mosquito species depend on season? Are there specific bacteria important in structuring the microbiota that dependent on regional environmental acquisition?

Some microbes, in particular the vertically transmitted endosymbiotic bacteria *Wolbachia*, have been shown to modulate pathogen infection and transmission in insects (e.g., Dennison et al., 2014; Dutra et al., 2016). *Wolbachia* endosymbionts affect the capacity of mosquitoes to carry specific parasites and viral pathogens (Martinez et al., 2014). *Wolbachia*-mediated effects in different hosts and RNA viruses range from reduced virus proliferation and transmission (Lu et al., 2012) to enhanced infection rates (Dodson et al., 2014). For instance, dengue virus can be suppressed by *Wolbachia* strains transinfected in *Aedes aegypti* (Hoffmann et al., 2011; Sinkins, 2013), and, at sufficiently high densities, in *Aedes albopictus* (Lu et al., 2012; Bian et al., 2013). In contrast, *Wolbachia* enhances WNV replication in *Ae. aegypti* cell line but inhibits virus assembly (Hussain et al., 2013), showing that *Wolbachia* protective phenotypes can rely on several distinct mechanisms. These mechanisms include resource competition (e.g., Moreira et al., 2009), immune stimulation (e.g., Pan et al., 2012), and small noncoding RNAs produced by *Wolbachia* that can regulate host genes (Mayoral et al., 2014). The protective effect of *Wolbachia* against Flaviviruses including Dengue and Zika (Dutra et al., 2016) has even been deployed deliberately for vector control. Artificially *Wolbachia*-infected mosquitoes were released in virus-endemic zones to spread the infection-reducing *Wolbachia* through the mosquito population (e.g., http://www.eliminatedengue.com).

Although artificially introduced *Wolbachia* strains can confer antiviral protection to new mosquito hosts (Bourtzis et al., 2014), similar effects have seldom been shown for native *Wolbachia* infections. For instance, while native *Wolbachia* infection in *Culex quinquefasciatus* inhibits dissemination and transmission of West Nile virus (WNV), the resistance is modest compared to the effects of *Wolbachia* in *Drosophila melanogaster* (Glaser and Meola, 2010). Natural resistance to WNV in field sampled *Cx. quinquefasciatus* and *Cx. pipiens* depends on sufficiently high *Wolbachia* densities, and is likely limited to specific populations (Glaser and Meola, 2010). In contrast to protection conferred by introduced *Wolbachia* strains, co-evolution of *Cx. pipiens* with natural *Wolbachia* infection favored vector competence and transmission of *Plasmodium relictum* (Zélé et al., 2014). *Wolbachia* may increase mosquito longevity and protect against *Plasmodium*-induced mortality (Zélé et al., 2014).

Wolbachia symbionts, though of unquestionable importance, are just one constituent of the entire mosquito-associated microbiota. Arguably, intracellular bacteria may not be considered part of the microbiota as they may have limited interactions with microbial communities in the mouth, gut, skin, or other organs. Because the gut epithelial cells are the initial site of viral proliferation, gut microbiota may play a crucial role in antiviral resistance and vector competence of mosquito species or populations (Moreira et al., 2009). One field of thought is that rather than stemming from co-evolution, the microbiota in mosquitoes might represent opportunistic environmental colonization (Osei-Poku et al., 2012). Undefined local processes were found to underlie spatial and temporal variation in vector competence for WNV in *Cx. pipiens* and *Cx. restuans* (Kilpatrick et al., 2010), and these results might be explained by location-specific environmentally acquired microbes.

To resolve these issues, we examined the microbiota, including *Wolbachia* relative abundance, in respect to host taxa, seasonality and WNV infection status in natural populations of 11 mosquito species in Ontario, Canada. Specifically, we tested whether the dominant drivers of microbial community variation were host species, geography, or season. We also tested whether any of the variation correlated with WNV infection, providing insight into possible effects on vector competence.

MATERIALS AND METHODS

Sample Origin, RNA, and DNA Extraction

Adult female mosquitos of 11 species were collected between 2011 and 2013 from Toronto, and 9 different geographical regions in Ontario, Canada (**Table 1**, **Figure 1**). Traps were set at residential properties, and at municipal buildings or parks. Details of sampling design and methods are available in Supplemental Materials (Table S1, Figure S1). The collected insects were frozen, identified morphologically to species, and pooled from each trap into samples containing 1–50 mosquitoes of the same species (**Table 1**). If only one individual of a species

TABLE 1 | Mosquito species captured and sampled for West Nile virus (WNV) from 2011 to 2013 in Ontario, Canada.

Species	N samples	Mean N mosquitoes/sample	N samples tested for WNV	WNV+ samples	WNV prevalence (%)
Aedes vexans complex	2,947	14.3	1,868	15	0.8
Ochlerotatus canadensis	201	4.1	19	0	0.0
Ochlerotatus japonicus	1,049	5.5	1,037	1	0.1
Ochlerotatus stimulans	143	9.3	143	0	0.0
Ochlerotatus triseriatus	459	4.1	458	1	0.2
Ochlerotatus trivitatus	472	8.2	472	0	0.0
Coquilletidia perturbans	1,139	16.3	1	0	0.0
Culex pipiens/restuans	3,652	12.2	3,648	297	8.1
Anopheles punctipennis	406	2.4	406	1	0.2
Anopheles quadrimaculatus	70	2.3	69	0	0.0
Culex salinarius	109	3.4	109	4	3.7
Culex tarsalis	2	2.5	2	0	0.0
Total	10,649	11.3	8,232	319	3.9

was present in a trap, this provided an unpooled sample with only one mosquito.

Following homogenization and centrifugation, RNA was extracted from 200 μL of supernatant (Supplemental Methods). RNA from pools found to be positive by the WN3′ NC primer-probe combination was re-tested using the WNENV primer-probe combination to confirm WNV positivity. The Ontario Ministry of Health mandated that this protocol be utilized in the mosquito surveillance program for the testing of WNV in mosquito pools. The amplicon sizes are 103 bp for the WN3′NC primers and 70 bp for the WNENV primers as previously reported (Lanciotti et al., 2000).

From each sample, DNA was extracted according to the Earth Microbiome Project protocol (http://www.earthmicrobiome.org/emp-standard-protocols/dna-extraction-protocol/) using the MoBio PowerSoil DNA Isolation Kit for 2,298 samples (232 single mosquito isolates and 2,066 pooled samples representing gDNA from 2 to 50 individuals of the same species).

WNV Diagnosis

A total of 8,232 samples were tested for WNV between 2011 and 2013 using a TaqMan real time PCR assay according to Lanciotti et al. (2000; Supplemental Methods and Table S2).

Data Generation and Processing

Genomic DNA from 2,298 samples, along with the negative controls, was amplified according to the EMP protocol (http://www.earthmicrobiome.org/protocols-and-standards/16s/) using the 515/806 primer pair and analyzed using barcoded sequencing on 3 lanes of Illumina HiSeq 2000. Raw reads of 125 bp were processed using UPARSE (Edgar, 2013) according to the following scheme: (i) demultiplexing reflecting the raw data barcodes, (ii) quality filtering using the maxee parameter set to 0.5, (iii) dereplicating identical sequences, and (iv) removing singletons to create *de novo* database (v) mapping raw reads to the database to generate sequence counts per OTU and sample. The number of sequences per sample was approximately 50,000 on average, and samples with <1,000 reads were omitted. While the number of the OTUs appeared extremely high in comparison with previous studies (e.g., Coon et al., 2014), the raw data were reanalyzed using Deblur (https://github.com/biocore/deblur; AA, submitted). This algorithm does not inflate the number of OTUs and produced a more realistic picture of the mosquito microbiota. Deblur is a de-noising method which, after removal of PCR and read-error derived reads, can identify sequences with as little as one nucleotide difference over the sequence region, as opposed to clustering based approaches such as UPARSE, which cluster together sequences more similar than a given noise derived threshold (usually 97%). For instance, for the 11 mosquito species analyzed, deblurring revealed 17 *Wolbachia* clusters (called OTUs in a broader sense here after, Figure S3) compared to 148 OTUs generated by UPARSE. Another useful property of Deblur is that it is stable—it is run on each sample independently and the same sequence in different samples will be identified as the same OTU. Whereas, in de-novo clustering methods, all samples need to be processed together since otherwise the same sequence can be assigned to two different OTUs, depending on the neighboring sequences.

After sequence processing, negative controls were checked for contaminants. Two out of six negative controls were clean (<30 sequences), while the other four showed some amplification (above 5,000 sequences). Particularly, most of the sequences in those negative controls corresponded to two OTUs, identified as Enterobacteriaceae (54% of the reads in one control) and Pseudomonadaceae (between 43 and 63% of the reads in three controls). The Pseudomonadaceae OTU showed high frequency (90%) and mean abundance (269.4 ± 230.7 sequences, normalized at 1,000 sequences per sample) in our samples. This result agrees with previous work that found this family of bacteria in different mosquito species (Minard et al., 2013). Accordingly, the amplification in these controls could be caused by cross-contamination from our samples during processing, more likely than from an external source. On the other hand, the Enterobacteriaceae OTU showed low frequency (20%) and mean abundance (19 ± 70.2 sequences per sample, normalized at 1,000 sequences) in our samples, suggesting that it could be a real contaminant. However, the presence of insect-specific symbionts such as *Wolbachia* in high abundance in our samples

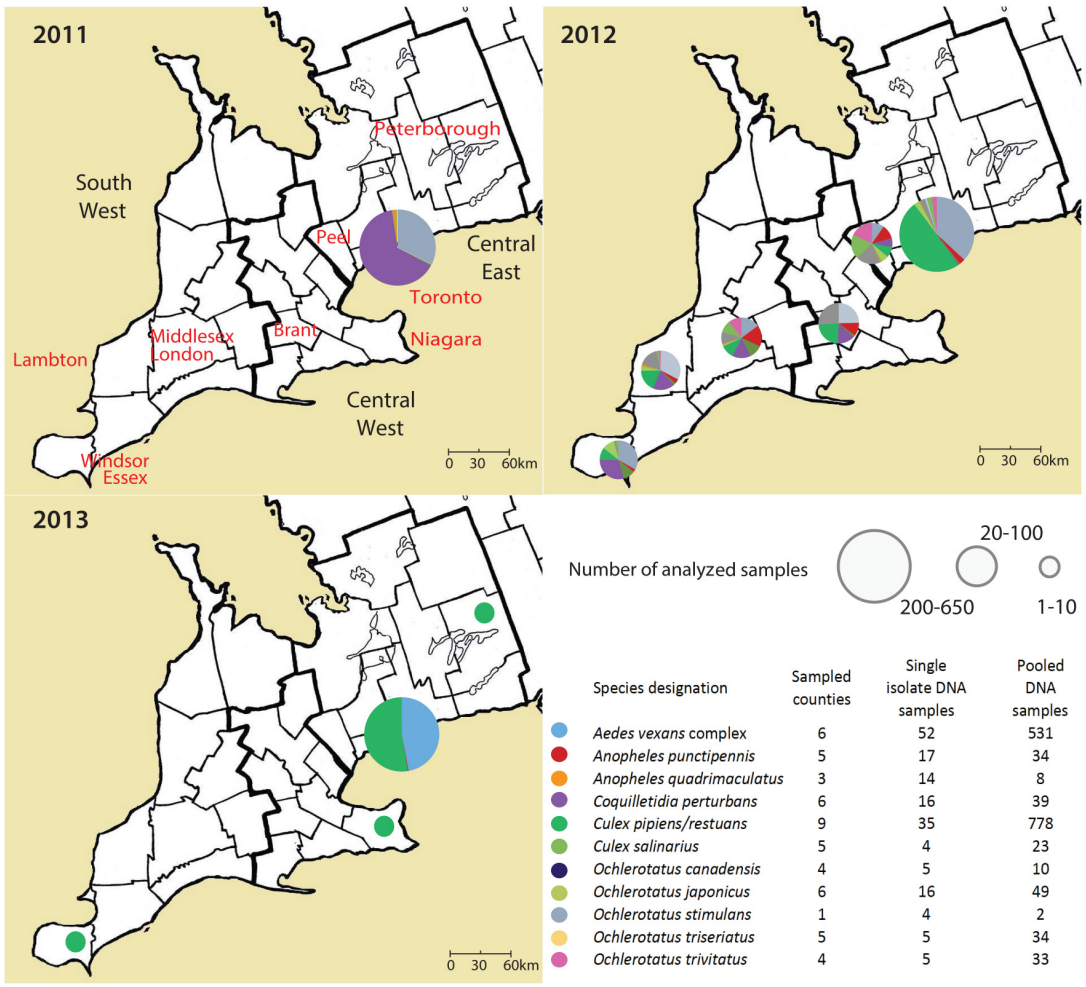


FIGURE 1 | An overview on analyzed samples, their geographical background and year of sampling in Ontario, Canada. Single mosquito samples and pooled mosquito samples were analyzed separately.

(but not in the negative controls), and its mosquito species-specific pattern, suggest that there was no significant effect of any contamination in our samples that would affect further analyses.

Clustered OTUs consisted of sequences matching bacterial and mitochondrial 16S rRNA genes as well as 18S rRNA gene sequences that were also amplified (presumably because of low primer specificity and low complexity of the analyzed microbiota against an excess of host DNA). 16S rRNA OTUs for analyses of microbiota were retrieved from the complete data set using BlastN searches against 16S rRNA gene sequence database (NCBI). The taxonomic assignment of these OTUs was based on the RDP classifier and Greengenes reference using 97% similarity (Wang et al., 2007). Considering recent findings that 16S rRNA amplicon sequencing can reveal relative quantitative changes in abundance of taxa among samples (D'Amore et al., 2016), the relative abundance of *Wolbachia* in the single isolates was calculated as the percentage of all the 16S rRNA amplicon reads. Other sequences (18 and 16S mitochondrial, plastid and archaeal OTUs) were identified using BlastN (Camacho et al., 2008). This approach enabled a strict quality check (discarding possible contaminants and taxonomically mis-assigned samples using 18S rRNA gene sequences described below).

Although the mosquito specimen identification was solely based on morphology, we took advantage of 18S rRNA amplicons and used those as a molecular marker. Indeed, we retrieved on average 1,118 and 4,376 reads of mosquito 18S rRNA per each of individual and pooled samples, respectively. The data were used as a quality check with the potential to reveal and resolve several methodological artifacts. In particular, artifacts could include incorrect taxonomic assignment based on morphology, species complexes that cannot easily be resolved, and sample contamination from other mosquitoes in the same trap. Clustering of 18S rRNA host sequences (detected here with

the universal 16S primers) displayed a clear pattern reflecting the sample taxonomy and allowed for molecular based taxonomic determination on different taxonomic levels. While we could not distinguish between closely related species of *Aedes vexans* complex, or among four *Ochlerotatus* species (*O. canadiensis*, *O. stimulans*, *O. triseriatus*, and *O. trivitatus,* all clustered into a single OTU), *O. japonicus* sequences formed another OTU. *Anopheles* species, *An. punctipenis* and *An. quadrimaculatus*, split into two different OTUs with 98% simillarity. Two *Culex* species, *Cx. pipiens/restuans* and *Cx. salinarius* however clustered together into a single OTU. *Coquilletidia perturbans* sequences were represented by a unique cluster.

The following rule was applied to filter out potentially misleading data: Samples with <90% of 18S rDNA sequences in the taxon specific OTU described above, and samples with 0 total reads for host 18S rRNA. These samples were not analyzed within the final dataset. Altogether 102 pooled samples and 21 individual samples were discarded. Taxonomic assignment was corrected for eight samples. Altogether, a subset of 173 single-mosquito samples, and 1,541 pooled samples (2–50 mosquitoes of the same species trapped together) passed the quality control and was further analyzed (**Figure 1**). The raw sequence data are available at European Bioinformatics Institute database under accession number ERP021438. The dataset is also available at https://qiita.ucsd.edu/ (ID 10815).

Statistical Analyses of the Microbial Communities

To assess composition and diversity of mosquito associated bacterial communities, two sets of 16S rDNA amplicons were analyzed: single-mosquito samples and pooled samples, following the same workflow. All the analyses were performed in R environment (R Core Team, 2016) using following packages and libraries: datasets, dplyr, stats, biom, vegan, ggplot2, clickme (Wickham, 2009; Oksanen et al., 2013; McMurdie and the biom-format team, 2014; RStudio Team, 2015; Caballero, 2016; R Core Team, 2016). First, the sequencing depth among the samples was normalized by rarifying the data to 1,000 sequences per sample for the single-mosquito samples, and 5,000 sequences per sample for the pooled samples. Normalization of sampling depth is advised for samples ranging widely in sequencing depth (Weiss et al., 2015). Shannon index and richness was used to describe the bacterial diversity among different host species. Kruskal-Wallis tests were used to evaluate differences in diversity among host species. Bray-Curtis dissimilarities calculated from abundance tables were used for further evaluation of selected factors, i.e., host genetic background, geographical background, seasonality (week number), potentially shaping the community profiles. Statistical testing was performed using permutational multivariate ANOVA implemented in R (Adonis function in vegan package; Oksanen et al., 2013). In order to reveal to what extent *Wolbachia* OTUs affect calculated dissimilarities, these OTUs were systematically excluded generating a series of datasets (not shown). The dissimilarities among analyzed microbiomes were then statistically tested as described above for the host species and genus level. Furthermore, a pairwise comparison was performed for each possible species and genera pair for the full dataset and the one missing all the *Wolbachia* OTUs. Constrained ordinations were used to visualize the overall differences among microbiomes of different species and genera. The two control analyses for the exclusion approach were performed eliminating the second most abundant OTU, i.e., *Asaia*, and *Pseudomonas,* the OTU shared by all the mosquito species. All the datasets used in the exclusion analyses underwent rarefaction at the level of 300 reads acceptable for the majority of the samples. QIIME implemented python script *group_significance* was used with Kruskal-Wallis tests to identify bacterial OTUs with significantly different abundances among species.

Phylosymbiosis Analysis

Phylosymbiosis refers to the observation of congruency between host phylogeny and whole microbial community topology, and infers some shared ancestral microbial community. Using the same analysis as presented in Brooks et al. (2016), host phylogenetic trees were constructed using an incomplete multigene matrix of 18S, 28S, COI and NADH available for all analyzed mosquito species in GenBank. The sequences were aligned using Muscle v3.8.31 (Edgar, 2004), and alignments were evaluated using jModelTest v2.1.7 (Darriba et al., 2012). The optimal host tree and bootstrap values were generated in RaxML v8.0.0 (Stamatakis, 2014). The software package ETE 3 (Robinson and Foulds, 1981) was used to determine topological congruencies for the host phylogeny and the beta-diversity of the average community abundancies for each host species. Topographical symmetry and edge similarity for trees was quantified by the normalized Robinson-Foulds (RF) metric (Huerta-Cepas et al., 2016) to determine topological similarity on a scale from 0 (complete congruence) to 1 (incomplete incongruence). Robinson-Foulds metrics were evaluated for Bray-Curtis, unweighted UniFrac, and weighted UniFrac beta-diversity dendrograms at 97% and 99% OTU clustering and compared to a null to determine if the host-microbe congruency is randomly associated (Brooks et al., 2016).

RESULTS

Diversity and Host Species Specificity of Mosquito Microbiota

Single-mosquito samples were different than pooled samples (noted hereafter as ** for 95% confidence and *** for 99% confidence) in diversity of mosquito microbiota (Mann-Whitney *U*-test for richness: $U = 107{,}639^{***}$; Mann-Whitney *U*-test for Shannon diversity index: $U = 144{,}615^{**}$). The average total read number for individuals was 42,990, and 82,124 for pooled samples. The mean (SD) bacterial richness in microbiota of all 11 species was 50.9 (17.3) bacterial OTUs for single-mosquito samples and 64.8 (31.3) bacterial OTUs for pooled samples (**Figure 2**). Highly abundant taxa found to be associated with at least one of the analyzed mosquito species were primarily of the phylum Proteobacteria, including *Asaia*, *Wolbachia*, *Serratia*, *Pseudomonas* and other bacteria from the family Enterobacteriaceae. Except for the Proteobacteria, members of Entomoplasmatales (Tenericutes) were also found in

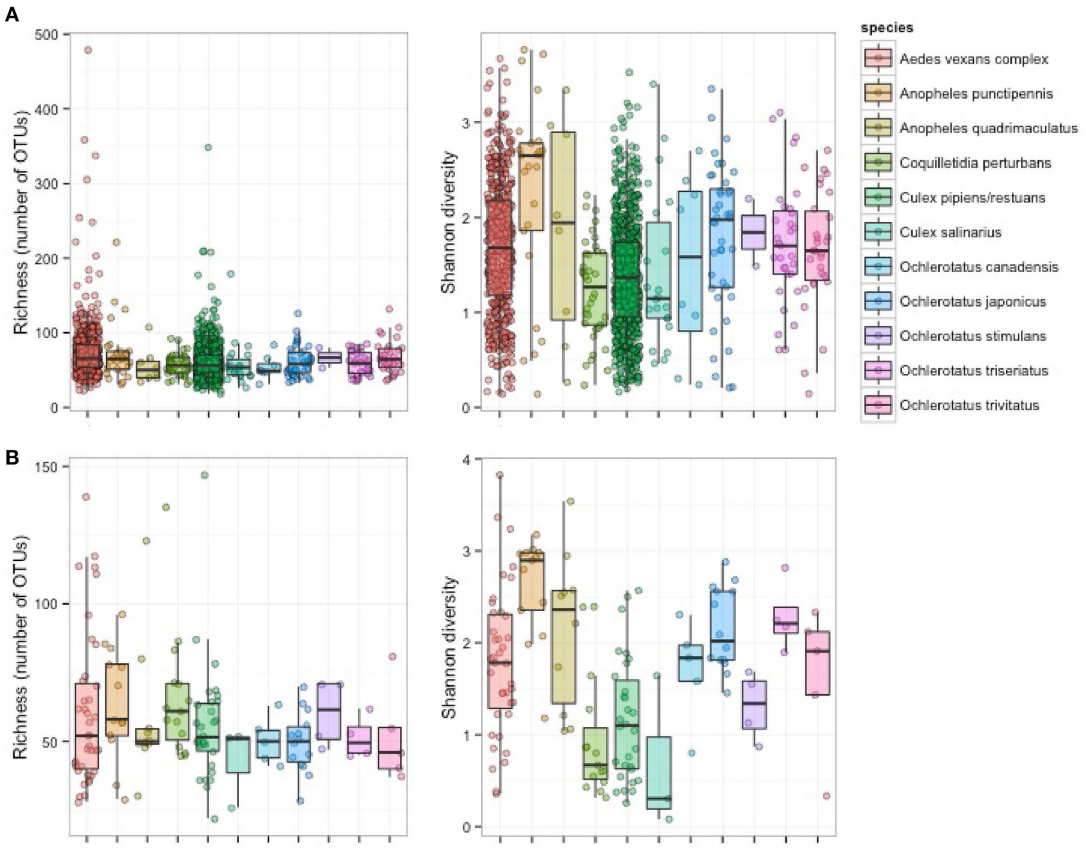


FIGURE 2 | Richness and alpha diversity of microbiota from 11 mosquito species based on pooled samples (A) and single mosquito samples **(B)**.

high numbers in some *Ochlerotatus* species. Relative abundances of these principal bacterial taxa calculated for single-mosquito samples are shown in **Figure 3** (Figure S2 pooled samples). There were 36 OTUs that differed significantly among species based on Kruskal-Wallis test with FDR correction (Supplemental data: group_significance_results.xlxs). There were significant differences among the microbiota of different host species using Bray-Curtis dissimilarity matrices, and differences were less distinct after removal of *Wolbachia* symbionts (**Figures 4**, **5**).

Phylosymbiosis Analysis

All beta-diversity distance matrices indicated an accurate separation of the *Anopheles* genus, and some conservation of phylosymbiosis between major genera is maintained when average bacterial communities are clustered at 97 or 99% OTU identity (**Figure 6**). The relationship of the host phylogeny and the 97% OTU clustering of microbial communities is nearly completely incongruent with the exception of the weighted unifrac (RF index of 0.75, Table S3). However, as recently observed in Brooks et al. (2016), when microbial communities are clustered at 99% OTU identity, all beta-diversity analyses conducted indicate significant phylosymbiosis for the wild mosquito species and their respective microbial communities (Table S3).

Wolbachia

Within the 11 species analyzed, three (*Cx. pipiens/restuans*, *Cx. salinarius* and *Cq. perturbans*) were found to harbor *Wolbachia* in high numbers (**Figure 4**). In *Cx. pipiens* pooled samples (n = 591), *Wolbachia* was not detected in one sample (a 6 individuals pool sampled in June 2011), indicating a high prevalence in this mosquito species. A total of 17 *Wolbachia* OTUs (reduced to 13 by phylogenetic analysis, Figure S3) were found among the 11 mosquito species, but only 3 were found at high abundance. Ten species were associated with *Wolbachia* OTU1 or a mixed infection, *Coquilletidia perturbans* harbored a distinct *Wolbachia* strain (represented by *Wolbachia* OTU2 with 94.2% similarity in 125 bp to OTU1, **Figure 4**, Figure S3). *Wolbachia* symbionts dominated microbial communities of *Culex* and *Coquilletidia* species profiled in **Figures 5A,C**. Comparing microbiota at the host genus level with pairwise comparisons revealed significant differences between all the pairs, except for the *Aedes-Anopheles* pair (**Figure 5A**; bold underlined numbers stand for Adonis R^2-values significant at the 99% level). In contrast, using the filtered dataset lacking all *Wolbachia* OTUs, no significant differences

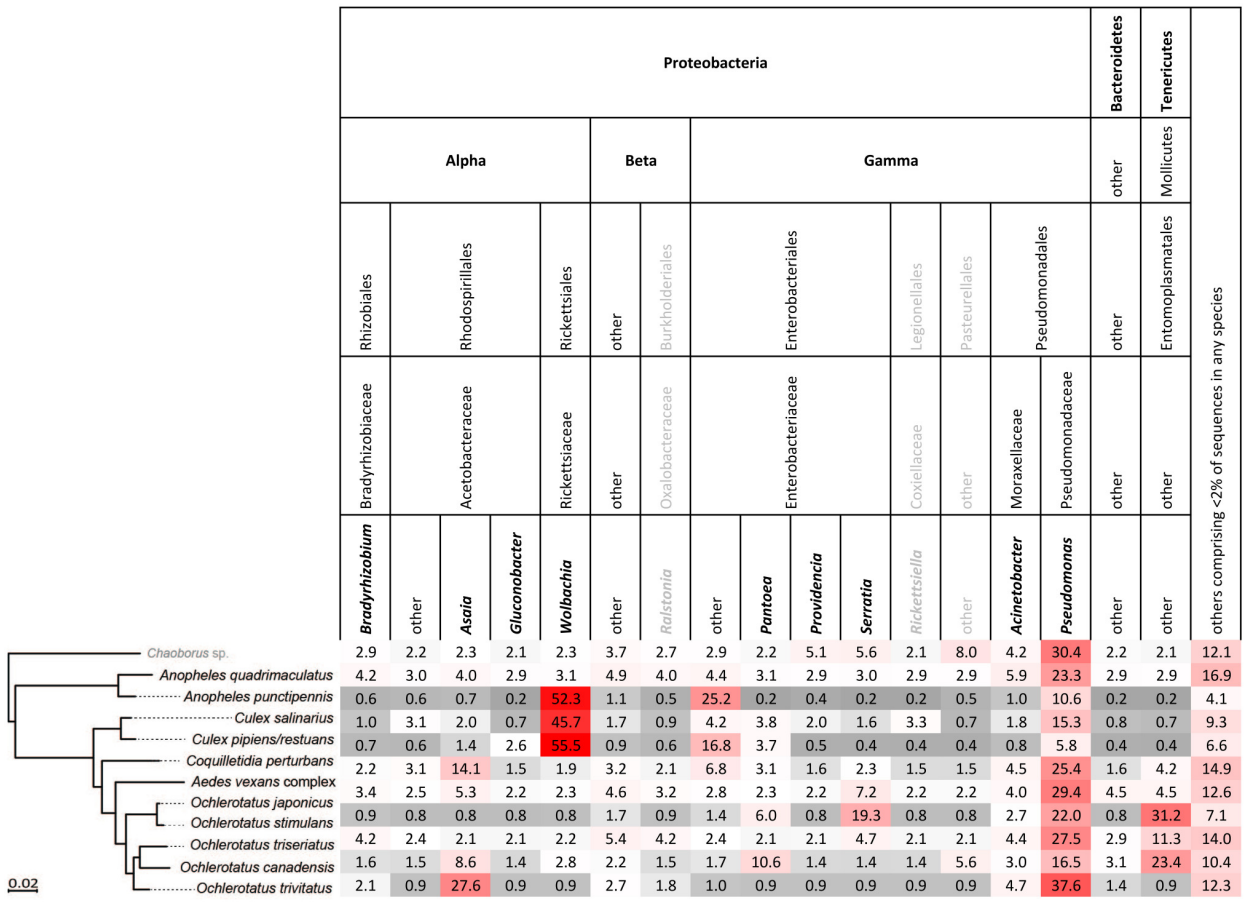


FIGURE 3 | Relative abundance of bacterial taxa in each host species arranged in order of phylogenetic relationship, based on single mosquito samples. Bold printed taxa were congruently found in abundances above 2% for the pooled samples (Figure S2).

were found among the mosquito genera pairs (**Figure 5**). Statistical evaluation of differences between mosquito species pairs for complete and *Wolbachia* free datasets is provided in Table S4, highlighting the distinctive effect posed particularly on the microbiome profiles of *Cx. pipiens/restuans* and *Coquilletidia perturbans* by these bacteria.

Biogeography and Seasonal Effects on Mosquito Microbiota

The effect of geographical background on the microbiota was tested for pooled samples of *Ae. vexans* complex and *Cx. pipiens/restuans*, the two taxa with sufficient sample sizes for statistical evaluation. Because we did not find significant differences among years, tests for biogeographical effect were performed over the 3-year period. Site of capture did not significantly affect microbiota of *Ae. vexans* complex pooled samples from six different regions (**Figure 1**; Adonis: R^2 = 0.00845). Similarly, site was not a significant factor differentiating microbiota from single *Ae. vexans* complex samples from Brant and Toronto (R^2 = 0.0659). Pooled samples from *Cx. pipiens/restuans* allowed for testing among nine sites. The analyses produced significant results for differences between following regions: Brant-Windsor Essex (R^2 = 0.12065***), Peterborough- Windsor Essex (R^2 = 0.2948***) and Haldimand-Windsor Essex (R^2 = 0.17541***). Windsor Essex was the most distinct site and at the edge of the sampling region (**Figure 1**).

Along with seasonality in mosquito density (Table S1, Figure S1), we found overall seasonal fluctuations in the microbiota in *Ae. vexans* complex and *Cx. pipiens/restuans* (**Figure 7**, Table S5, Figures S4, S5). The OTUs with the greatest seasonal dynamics (largest effect sizes) are indicated in Table S5, with trends illustrated in heatmaps for each species (Figures S6, S7). The abundance of *Wolbachia* shifted seasonally in *Cx. pipiens/restuans* with a dip in June-July, but not in *Ae. vexans* complex (**Figure 6E**). Four other OTUs showed seasonal trends in each of the 3 years sampled both in *Cx. pipiens/restuans* and in *Ae. vexans* complex including Acetobacteraceae, Bacteroidetes,

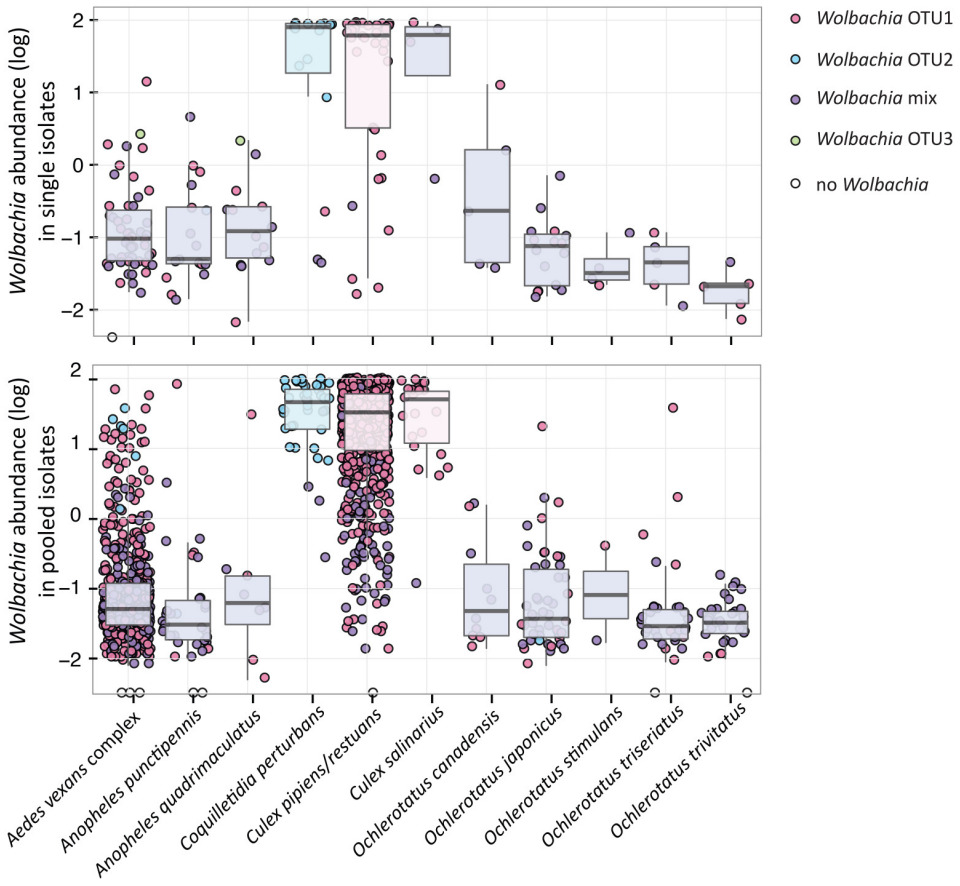


FIGURE 4 | *Wolbachia* relative abundance and presence of dominating OTUs for 11 analyzed species sampled as individuals and pools. The relative abundance represents 16S rRNA read percentage assigned to any *Wolbachia* OTU. The color key for individual points (samples) and box plots reflects the presence of particular OTU(s), i.e., 100% of reads assigned to the single *Wolbachia* OTU in red, blue and green; mixed infection in purple.

Enterobacteriaceae, and *Asaia* (**Figure 6**). Seasonal dynamics were not analyzed for the 9 other host species with smaller sample sizes.

Seasonal Changes in WNV Prevalence and Microbiota

Sampling mosquitoes for WNV and microbiota across a 3 year period in Ontario revealed 6 species as potential vectors for WNV (**Table 1**), with the highest prevalence in the *Cx. pipiens/restuans* and a seasonal spike in prevalence each year in early to mid August reaching up to 43% of pooled samples (**Figure 7F**). Species exhibiting low relative abundance of *Wolbachia*, including all *Ochlerotatus*, *Aedes* and *Anopheles* specimens, were identified as potential WNV carriers. Samples of *C. perturbans* associated with *Wolbachia* OTU2 in high densities and were found in other studies to have low WNV infection prevalence (Sardelis et al., 2001; Cupp et al., 2007). Out of seven species with single mosquito samples showing some WNV positives, 6 species had higher mean abundance of *Asaia*, and 7 species had higher mean abundance of *Wolbachia* in WNV uninfected compared to infected mosquitoes. In *Cx. pipiens/restuans* samples, mean *Wolbachia* reads were approximately 68.8% ($N = 31$) in WNV negative samples compared to 0.3% ($N = 3$) in WNV positive samples. There was a dip in *Wolbachia* prevalence and a nearly corresponding spike in WNV prevalence in pooled samples of *Cx. pipiens/restuans* (**Figure 7**). Conditions in the weeks prior to sampling were critical in driving patterns of WNV.

Wolbachia abundance in *Cx. pipiens/restuans* pooled samples negatively correlated with temperature (**Figure 8A**). There was a significant correlation between *Wolbachia* abundance 3 weeks before sampling and WNV prevalence ($R^2 = 0.42249$, $P = 0.012$, **Figure 8B**). The correlation coefficient increased with time prior to sampling for WNV prevalence vs. temperature and WNV vs. *Wolbachia* abundance; precipitation did not correlate with WNV prevalence (Table S6). Temperature negatively correlated with *Wolbachia* abundance (**Figure 8A**), and temperature 3–4 weeks prior to sampling correlated with WNV prevalence (Table S6). Thus, higher temperatures may have led to decreased *Wolbachia*; through vertical transmission to the subsequent generation, reduced *Wolbachia* is hypothesized to increase susceptibility to WNV (**Figure 8C**). Given that a 2°C increase in peak summer

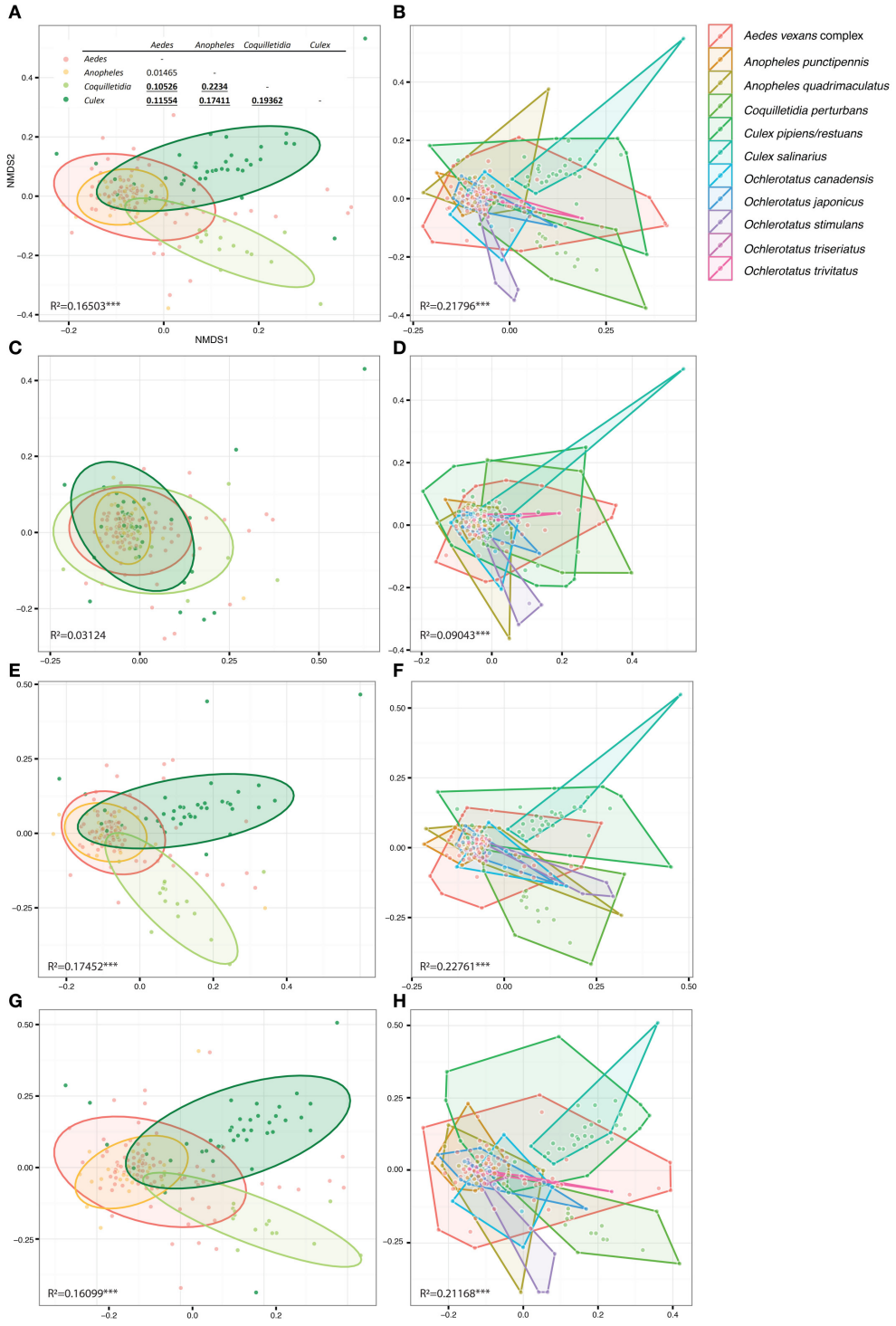


FIGURE 5 | Overall differences among microbiomes of different mosquito genera (right) and species (left) plotted in constrained ordinations. (A,B) were produced using the complete single mosquito dataset; **(C,D)** are based on data from which all *Wolbachia* OTUs were removed in order to test *Wolbachia* effects on dissimilarity among microbiota profiles (for more details see Materials and Methods). **(E,F)** present control analyses excluding the next most abundant OTU, i.e., *Asaia* **(E,F)**, from the data set, and an OTU shared by all the mosquito taxa, i.e., *Pseudomonas* **(G,H). (A)** includes pairwise statistical evaluation: bold underlined numbers stand for R^2-values significant at 99% confidence interval calculated for dissimilarities of genera pairs. R^2−values indicate statistical evaluation of dissimilarities among all genera/species in each plot. Considering the low number of samples per species, hulls were used to highlight the corresponding points, instead of the statistical ellipses used for genera based analyses.

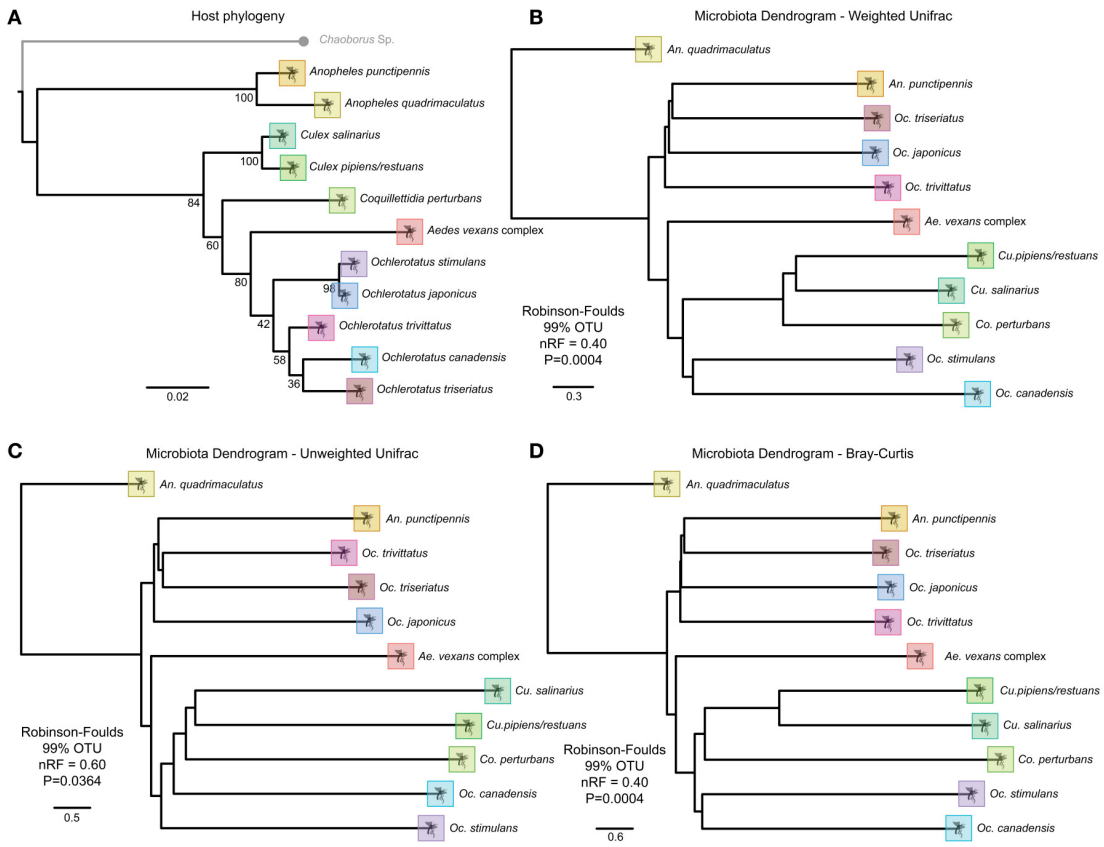


FIGURE 6 | Topographic trees of host phylogeny and OTU beta-diversity metrics indicating strong phylosymbiosis of mosquito host species and microbiota. (A) RaxML host phylogeny based on an incomplete multigene matrix of 18S, 28S, COI and NADH. Beta-diversity analysis for each host species at 99% OTU clustering for **(B)** Weighted UniFrac, **(C)** Unweighted UniFrac, and **(D)** Bray Curtis metrics.

temperature would decrease *Wolbachia* abundance by 22% (**Figure 8A**), this reduction of *Wolbachia* could lead to an 18% increase in WNV prevalence from 4.7 to 5.5% of samples positive (**Figure 8B**). This scenario of climate change is realistic in eastern North America, particularly in urban areas (Primack, 2014). We suggest that WNV prevalence in *C. pipiens/restuans* may increase in samples collected after sampling for this study completed in 2013.

DISCUSSION

Recent microbiome studies focus on factors driving the composition and function of host microbiota. In this study, we examined 11 adult mosquito host species from six regions in southern Ontario, Canada. Mosquitoes were sampled over 3 years in the Toronto region. We found that host species was the largest driver of the microbiota, while region had little impact for the species tested (*Cx. pipiens/restuans* and *Ae. vexans*). However, the region with the most distinct microbiota, Windsor Essex, was at the edge of the sampling region, indicating that over larger geographical scales than studied here, region may be an important factor driving microbiomes, or that region is correlated with important environmental conditions. Seasonal shifts were consistently repeated over the 3-year period in microbiomes of *Cx. pipiens/restuans* and *Ae. vexans* complex. Both host species and seasonal shifts in microbiota correlate with patterns of WNV in these mosquitoes.

In accordance with previously published results on *Anopheles* and *Culex* genera (e.g., Gimonneau et al., 2014; Duguma et al., 2015), we found that microbiota of *Aedes*, *Ochlerotatus*, *Anopheles, Culex,* and *Coquilletidia* species were dominated by the phylum Proteobacteria. This common pattern suggests that some characteristics of the Proteobacteria may make them especially suitable for mosquito colonization. Interestingly, although a clear environmental influence from the water stages to the adults has been detected (Coon et al., 2014; Tchioffo et al., 2016), the microbiota seems to differ specifically between mosquito genera or even between species within the same genus (e.g., Muturi et al., 2016) regardless their origin, suggesting a certain level of selection toward a beneficial microbiota (Gimonneau et al., 2014). Indeed, several predominant Proteobacteria have been found to have protective affects on mosquitoes including *Serratia* (Bando et al., 2013; Tchioffo et al., 2016) and *Wolbachia* (Moreira et al., 2009).

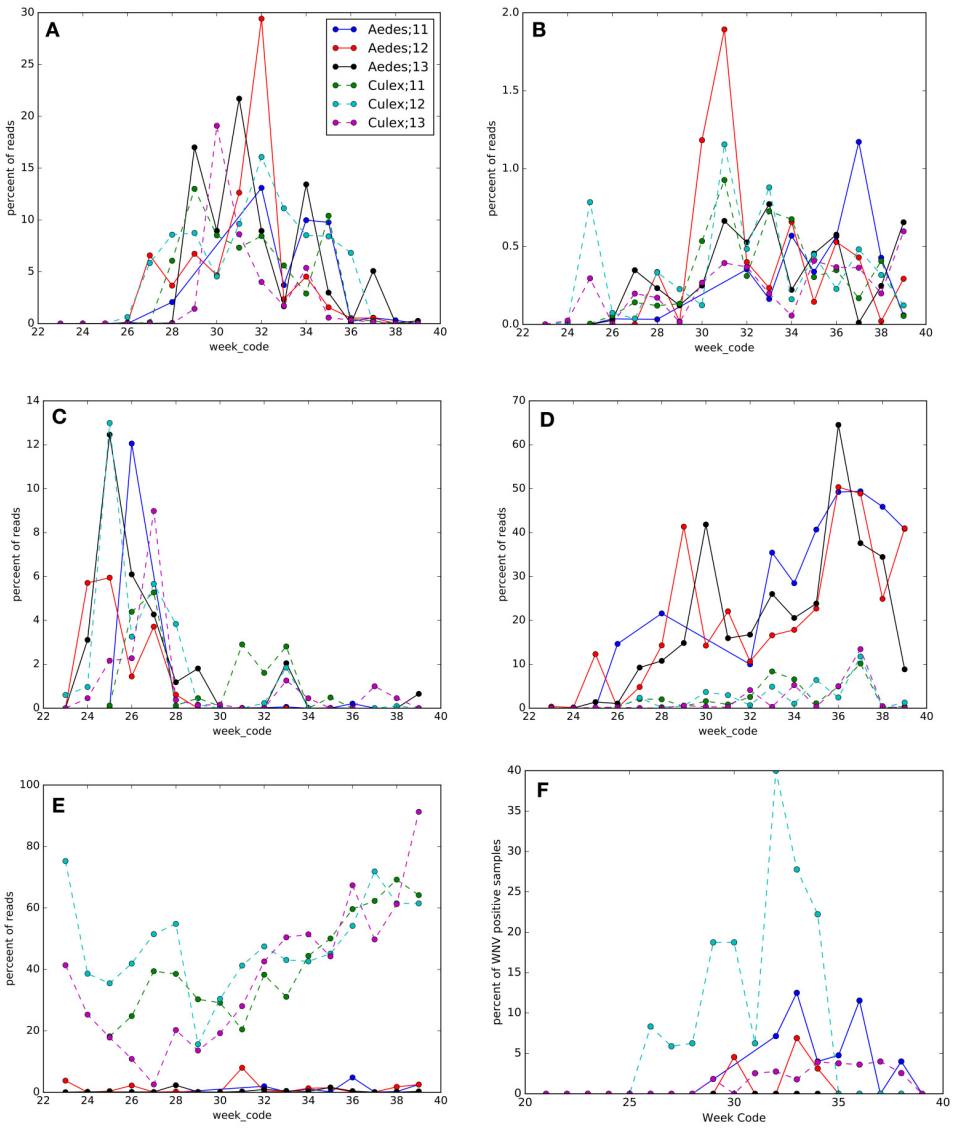


FIGURE 7 | Seasonal changes in OTUs and West Nile virus (WNV) prevalence in *Culex pipiens/restuans* (Culex) and *Aedes vexans* complex (Aedes) pools sampled between 2011 and 2013 (see key for species and year). Taxonomy of OTUs in each panel: **(A)** Proteobacteria; Alphaproteobacteria; Rhodospirillales; Acetobacteraceae, **(B)** Bacteroidetes, **(C)** Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae, **(D)** Proteobacteria; Alphaproteobacteria; Rhodospirillales; Acetobacteraceae; *Asaia*, **(E)** Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; *Wolbachia*, and **(F)** WNV prevalence.

Abundant *Pseudomonas* are commonly found across mosquito species (Charan et al., 2013; Minard et al., 2013). The presence of at least some bacterial strains, regardless of their origin, may be essential for successful mosquito development (Chouaia et al., 2012; Coon et al., 2014), digestion, and fecundity (Gaio et al., 2011).

This is the first observation of wild caught mosquitoes exhibiting phylosymbiosis under natural conditions. In the recent study by Brooks et al. (2016), laboratory reared mosquito species were isolated in near identical conditions without access to natural microbial communities or *Wolbachia* infections. In the study presented here, we observed significant ($p < 0.005$) similarities between the host species phylogeny and the microbial community composition within a given species using the same analysis. This strengthens the hypothesis that there is host selection on the microbial communities between species that is independent of environmental factors.

The present study brings further insight into microbiota of less-studied mosquito genera, i.e., *Ochlerotatus* and *Coquilletidia*, and reveals significant differences among all analyzed species and genera. These differences were mainly driven by the presence and abundance of *Wolbachia,* the widespread intracellular symbiont

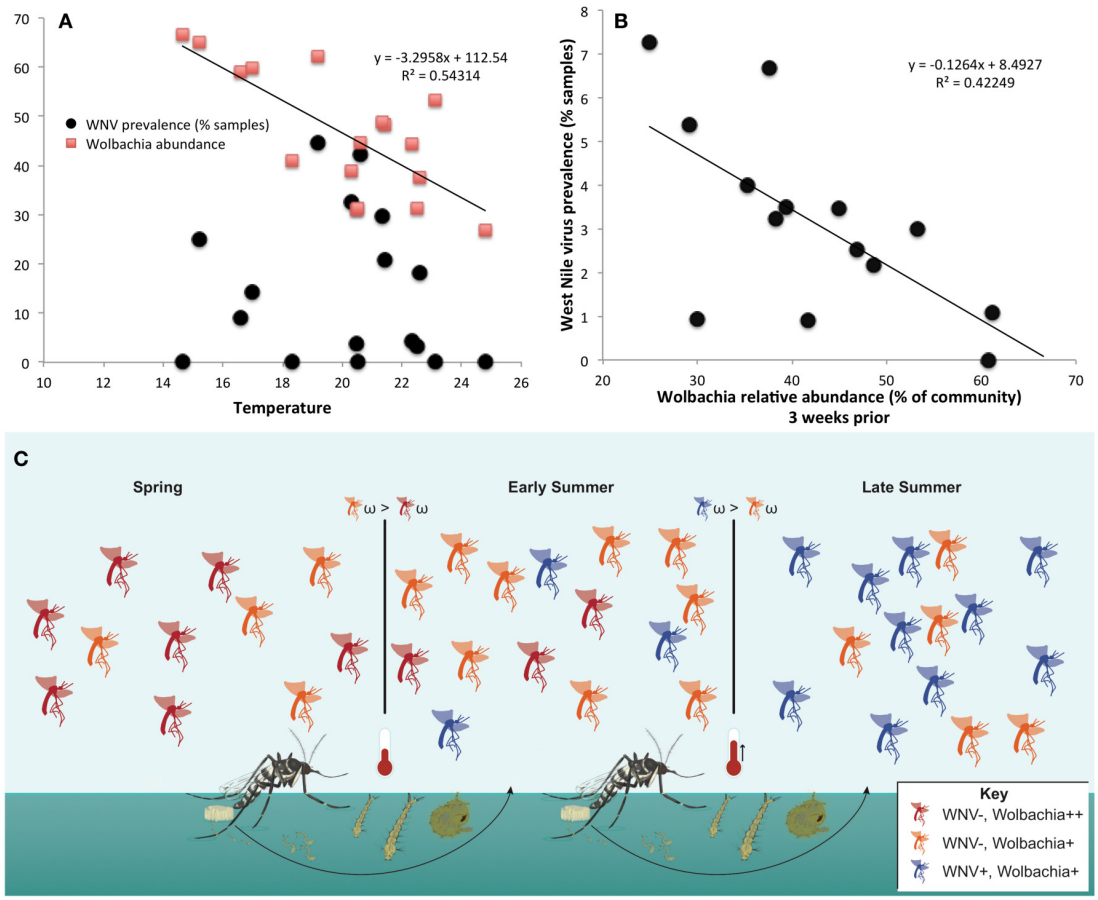


FIGURE 8 | Temperature negatively correlates with *Wolbachia* abundance, and *Wolbachia* abundance negatively correlates with West Nile virus (WNV) in *Culex pipiens/restuans*. (A) *Wolbachia*, but not WNV, is significantly correlated with mean weekly temperature in samples ($n = 538$) from Toronto. Daily climate data for 2011–2013 Toronto (WMO identifier 71265) downloaded from: http://climate.weather.gc.ca/historical_data/search_historic_data_e.html. **(B)** The relative abundance of Wolbachia in Cx. pipiens/restuans pooled samples ($n = 591$) over 3 years was significantly correlated with West Nile virus prevalence in pooled samples after a 3 week delay (see Table S6). Note that although the correlation coefficient increased from weeks 0 to 3, there was not a significant correlation until 3 weeks had elapsed. Thus, a decrease in Wolbachia abundance may contribute to conditions favoring WNV susceptibility in the succeeding mosquito generation. **(C)** A conceptual model illustrates the hypothesis that higher temperatures reduce Wolbachia abundance (orange) and lead to higher WNV prevalence (blue) in the following generation. Subsequently, fitness costs of WNV (ω), seasonal reductions in temperature, or lower density of mosquito larvae may drive the cyclical pattern and selection for increased abundance of protective Wolbachia (red).

with an immense diversity of strains and phenotypes (e.g., Werren et al., 2008). Indeed, two different strains of *Wolbachia* were found in high numbers in *Culex* and *Coquilletidia* species. *Wolbachia* dominance is particularly highlighted when removing individual *Wolbachia* OTUs from the analyses. Differences among and between most species/genera become insignificant when *Wolbachia* is excluded. While *Cx. pipiens/restuans* and *Wolbachia* wPip experienced a common evolutionary history (e.g., Atyame et al., 2011), *Wolbachia* symbionts may represent the selective force shaping the rest of the microbial community resulting in the exclusive characteristics of the entire system.

One recent study (Muturi et al., 2016) provided comparative results on microbiota of *Culex* species showing significant differences in relative abundance of dominant bacteria in *Cx. pipiens* and *Cx. restuans*. While our sample collection and analysis clustered together specimens in this morphologically indistinguishable species complex, their microbiota are more similar to those of *Cx. restuans* (56% Alphaproteobacteria and 21% Gammaproteobacteria), compared to *Cx. pipiens* (94% and 4% relative abundance, respectively) found by Muturi et al. (2016). In addition, Muturi et al. (2016) showed over 90% relative abundance of *Wolbachia* in *Cx. pipiens* from central Illinois compared to 47% in the species complex described here from southern Ontario.

Our results indicated four main bacterial genera dominating the analyzed microbiota, namely *Wolbachia, Asaia, Serratia* and *Pseudomonas*. Isolates of *Serratia* have been associated with anti-*Plasmodium* effects in some *Anopheles* mosquitoes (Bando et al., 2013; Tchioffo et al., 2016). *Wolbachia* and *Asaia* symbionts have also been previously described from other adult mosquitoes

(*Asaia* in different *Anopheles* species, Crotti et al., 2009; *Wolbachia* in several *Aedes* and *Culex* species, Sunish et al., 2011; Lu et al., 2012; Bian et al., 2013; Sinkins, 2013; Dutra et al., 2016; Muturi et al., 2016). These later two genera were demonstrated to be mutually exclusive, and in some mosquito species *Asaia* can prevent *Wolbachia* infection and vice versa (Rossi et al., 2015). Our results in suggest a similar trend of mutual exclusion in *Ae. vexans* complex and *O. trivitatus*. Additionally, the high abundance of *Pseudomonas* in *Anopheles, Ochlerotatus* and *Aedes* may also exclude the presence of *Wolbachia* in the system, suggesting that bacteria other than *Asaia* may affect the ability of the system to retain stable *Wolbachia* infection (Hughes et al., 2014). This is particularly relevant for disease transmission by mosquito vectors, as *Wolbachia* has been linked to vector competence (e.g., Micieli and Glaser, 2014).

In fact, a range of effects posed by *Wolbachia* on pathogens and parasites has been described for different insect hosts (Lu et al., 2012; Dodson et al., 2014). Particularly, some *Wolbachia* strains in combination with certain hosts are protective against viruses, but not others, as it happens in *Drosophila* (Osborne et al., 2009; Faria et al., 2016). In those cases of specific protective combinations, *Wolbachia* are found in higher densities compared to systems with non-protective *Wolbachia* phenotypes (i.e., in mosquitoes: Lu et al., 2012; Bian et al., 2013; and fruit flies: Osborne et al., 2009; Faria et al., 2016). These findings led some authors to the hypothesis that all *Wolbachia* strains are capable of antiviral protection if a sufficient density is reached, although that density level may be dependent on the host compatibility (e.g., Johnson, 2015). Here, with the presented data for *Cq. perturbans* and *Cx. pipiens/resturans* we question this general hypothesis. Both species, being capable of the WNV transmission (Sardelis et al., 2001), harbor different strains of *Wolbachia* in comparable abundances (median value calculated for single isolates are 78% for *Cq. perturbans* and 62% for *Cx. pipiens/restuans*, **Figure 4**). While 8.1% of 3,648 pooled *Cx. pipiens/restuans* samples were found positive for WNV, *Cq. perturbans* is not a priority species for WNV surveillance based on low field prevalence and low vector competence (Sardelis et al., 2001; Cupp et al., 2007). This further highlights the importance of host genetic background and *Wolbachia* strain combination, along with the symbiont abundance, as the main factors underlying host WNV carrier status, and vector competence. Our data, being generated from entire mosquito bodies, provide relative approximations on total *Wolbachia* numbers. The outcome of viral exposure may, however, depend on particular cellular or tissue levels of *Wolbachia* at the virus replication sites. Future research endeavors could thus combine the high-throughput population surveys with fluorescent *in situ* hybridization (FISH) approaches in order to localize and precisely quantify *Wolbachia* cells.

The microbiota of particular mosquito species may be an outcome of several factors. These include (i) host genetic background, (ii) long-term interactions among the bacteria and/or (iii) mutual interplay between host, microbiota and transmitted pathogen. This can be illustrated by recent findings of Martinez et al. (2015) on *Wolbachia* strain variation in terms of beneficial antiviral protection and parasitic cytoplasmatic incompability (CI). Strains that conferred antiviral effects negatively affected life-history traits and had a fitness cost compared to strains with CI in *Drosophila simulans*. Thus, persistence of antiviral *Wolbachia* strains in a mosquito population may depend on the prevalence and the burden of viral infections. Although few studies have examined burdens to mosquitoes of their vectored viruses, WNV caused increased mortality of *Cx. pipiens* and had strain-specific effects on fecundity and blood feeding behavior (Ciota et al., 2013).

In our data, we found a striking difference of WNV prevalence in two major vectors. Compared to an estimated 8.1% of WNV positive *Cx. pipiens/restuans* sample pools, the estimated WNV prevalence in populations of *Ae. vexans* complex is much lower (<1%). Low abundance of *Wolbachia* in this species indicates that other microbiota members, particularly *Asaia* and *Pseudomonas,* may confer antiviral protection with less fitness costs. Alternatively, apparently lower susceptibility to WNV infection may stem from host genetics, or differences in host feeding ecology.

Vector competence trials repeatedly demonstrate differences among strains of virus, and also among species and populations of vector mosquitoes. For example, *Ae. aegypti* from Santiago Island, Cape Verde exhibited high vector competence for DENV-2 and DENV-3 serotypes and a low susceptibility to DENV-1 and DENV-4 (da Moura et al., 2015). Variable population susceptibility to dengue virus has been attributed to differences in immune transcription (Carvalho-Leandro et al., 2012). Vector competence for the Asian genotype of Zika virus differed between populations of *Ae. aegygti* and between species *Ae. aegypti* and *Ae. albopictus* (Chouin-Carneiro et al., 2016). In addition to genetic differences among populations, we hypothesize that differences in vector competence are also caused by differences in microbiota affecting immune gene expression.

Wolbachia is one symbiont, among others, with known immuno-modulatory capacity in mosquitoes linked to vector competence (Kambris et al., 2009; Jupatanakul et al., 2014; Hegde et al., 2015). Here, we examined environmental effects on the microbiome as a potential mechanism for viral pathogen regulation, and found a striking correlation of season and temperature in particular that may regulate *Wolbachia* abundance in *Cx. pipiens/restuan*s hosts. *Wolbachia* abundance, in turn, may impact susceptibility to WNV infection status and prevalence of WNV at later time points. In experimental studies, as temperature was increased from 14 to 30°C, there was an increase in WNV titer in *Cx. tarsalis* (Reisen et al., 2006), indicating that climate can play an important role in disease dynamics. Temperature increases are known to reduce *Wolbachia* abundance across mosquito life stages (Wiwatanaratanabutr and Kittayapong, 2009; Ye et al., 2016). Ciota et al. (2014) examined life history traits of *Culex* mosquitoes. They found that days to emergence could range from approximately 25 to 12 days depending on temperatures of 16–24°C, respectively. Larvae with reduced *Wolbachia* could be sampled as adults as early as 2–3 weeks later depending on temperature, or adults with reduced *Wolbachia* could reproduce, and transmit a low abundance of *Wolbachia* to the next generation in that timeframe. Thus, reductions in protective microbiota mediated by climate warming in addition to increased

viral replication (Dohm et al., 2002) may lead to increased WNV and other arboviruses in both vertebrates and their mosquito disease vectors. Alternatively, independent of the seasonal changes in *Wolbachia* that are correlated with mean temperature, there may be an increase in WNV prevalence caused by a seasonal increase in infected blood-meal hosts.

A general role of seasonality has previously been suggested to affect microbial abundance in other blood sucking vectors including fleas and ticks (Lalzar et al., 2012; Cohen et al., 2015). In fleas, the spring-to-summer changes found in the bacterial community were attributed to the compositional changes in the diet, i.e., blood, including presence of pathogens (Cohen et al., 2015). The present study lacks the information on blood meal origin. However, considering seasonal fluctuations in bird populations, the preferred mosquito host and the reservoir for WNV, variation in the blood meal seems a plausible explanation for WNV seasonality. Additional environmental conditions may be responsible for seasonal effects detected in other microbiota. Some OTUs peak mid-summer such as Acetobacteraceae or Bacteriodetes, others such as Enterobacteriacea decrease yearly, or like *Asaia*, increase yearly, while other OTUs have fluctuating trends.

We found a strong seasonal pattern in WNV prevalence repeated over 3 years in *Cx. pipiens/restuans* mosquitoes. We also found seasonal patterns in other microbiota, indicating a potentially broad role for microbiota in pathogen defense and vector competence. This likely extends beyond *Wolbachia*, the dominant seasonal member in *C. pipines/restuans*, and current focus for disease mitigation against Flaviviruses (Dutra et al., 2016). Indeed, extended immunity provided directly by microbiota may be a trait under selection (Correa and Ballard, 2016; Faria et al., 2016), particularly if harboring pathogens has a fitness cost to the mosquitoes, as it does for WNV (Ciota et al., 2013). With increasing prevalence of mosquito-borne viruses there will be increased selection pressure on mosquitoes for symbiotic microbiota that increase resistance to viruses.

The most successful use of microbial management of insect vectors has been the application of *Bacillus thuringiensis* serotype *israelensis* (*Bti*) as a larvicide to reduce black fly populations in Western Africa to control onchocerciasis (Mbewe et al., 2014). *Bti* is now the only insecticide permitted in many European countries for mosquito control (Paris et al., 2011a) and *Bti* has become increasingly employed in mosquito control programs in the USA (Floore, 2006). Although resistance to other strains of *Bacillus thuringiensis* has been shown for several insect groups, the appearance of resistance to *Bti* toxins in natural vector populations has only recently been found in mosquitoes under some circumstances (Paris et al., 2011b; Bonin et al., 2015; Stalinski et al., 2014) but not in others (Araújo et al., 2013). The potential for such evolution presents a concern, and may be inevitable if *Bti* use becomes more prevalent. Like antibiotic resistance, a consistent use of chemical insecticides, including *Bti* toxins, sets the stage for selection in favor of resistant genotypes. We found that some *Bacillus* taxa increase seasonally (Table S5), and hypothesize that this may be influenced by applications of *Bacillus* larvicides and evolving resistance among *Cx. pipiens*. Indeed, an OTU matching a commonly used larvicidal agent, *Lysinibacillus sphaericus* (e.g., Valent Bioscience's VectoLex http://publichealth.valentbiosciences.com/products/vectolex), was found here on adult mosquitoes, although detection of genes involved in toxicity (i.e., Guidi et al., 2013) are needed to determine whether mosquitoes may be developing resistance to the larvicide.

CONCLUSION

The species analyzed here harbor significantly different microbial communities, all dominated by Proteobacteria. In this 3-year field survey we examined factors influencing the dynamics of mosquito microbiota. We found that host genetic background explained most of the variation, followed by season and geographic region as important drivers of the microbiome, similar to findings from other animal groups (e.g., Kueneman et al., 2014). Coevolution, and thus functional importance of the microbiome, is indicated by the relatedness of microbial communities of mosquito hosts in parallel to the host phylogeny (phylosymbiosis). A long-term coevolutionary relationship between *Wolbachia* and some host species may strongly influence the structure of the rest of the bacterial community. The presence of *Asaia* and *Pseudomonas* fluctuates with the presence of *Wolbachia* in mosquito hosts, supporting this hypothesis. The dynamic background of mosquito microbiota described here may help explain epidemiological patterns of WNV. For instance, if increasing temperatures cause a decrease in protective *Wolbachia*, climate warming may escalate disease caused by WNV. The importance of microbiota mediated by global change may have ramifications for mosquito-borne pathogens that are just beginning to be explored.

AUTHOR CONTRIBUTIONS

EN and DW contributed equally to this manuscript. DW, RK, JS, and EN contributed to the original idea; JS, AM, and DW contributed to field collection or lab analyses; DW, EN, SR, AA, RB, JL, JS, and AM contributed to data analyses. DW, EN, SR, RB, and RK wrote and revised the manuscript.

FUNDING

This project was partially funded by the Earth Microbiome Project, and the UMass Boston Proposal Development Grant Program to DW.

ACKNOWLEDGMENTS

We thank Todd Livdahl for helpful discussion, and Greg Humphrey and Gail Ackermann for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00526/full#supplementary-material

REFERENCES

Alto, B. W., Reiskind, M. H., and Lounibos, L. P. (2008). Size alters susceptibility of vectors to dengue virus infection and dissemination. *Am. J. Trop. Med. Hyg.* 79, 688–695.

Araújo, A. P., Araujo Diniz, D. F., Helvecio, E., de Barros, R. A., de Oliveira, C. M., Ayres, C. F., et al. (2013). The susceptibility of *Aedes aegypti* populations displaying temephos resistance to *Bacillus thuringiensis* israelensis: a basis for management. *Parasit. Vectors* 6:297. doi: 10.1186/1756-3305-6-297

Atyame, C. M., Delsuc, F., Pasteur, N., Weill, M., and Duron, O. (2011). Diversification of Wolbachia endosymbiont in the *Culex pipiens* mosquito. *Mol. Biol. Evol.* 28, 2761–2772. doi: 10.1093/molbev/msr083

Bando, H., Okado, K., Guelbeogo, W. M., Badolo, A., Aonuma, H., Nelson, B., et al. (2013). Intra-specific diversity of *Serratia marcescens* in Anopheles mosquito midgut defines Plasmodium transmission capacity. *Sci. Rep.* 3:1641. doi: 10.1038/srep01641

Bian, G., Zhou, G., Lu, P., and Xi, Z. (2013). Replacing a native Wolbachia with a novel strain results in an increase in endosymbiont load and resistance to dengue virus in a mosquito vector. *PLoS Negl. Trop. Dis.* 7:e2250. doi: 10.1371/journal.pntd.0002250

Bonin, A., Paris, M., Frérot, H., Bianco, E., Tetreau, G., and Després, L. (2015). The genetic architecture of a complex trait: resistance to multiple toxins produced by *Bacillus thuringiensis* israelensis in the dengue and yellow fever vector, the mosquito *Aedes aegypti*. *Infect. Genet. Evol.* 35, 204–213. doi: 10.1016/j.meegid.2015.07.034

Bonizzoni, M., Gasperi, G., Chen, X., and James, A. A. (2013). The invasive mosquito species *Aedes albopictus*: current knowledge and future perspectives. *Trends Parasitol.* 29, 460–468. doi: 10.1016/j.pt.2013.07.003

Bordenstein, S. R., and Theis, K. R. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol.* 13:e1002226. doi: 10.1371/journal.pbio.1002226

Bourtzis, K., Dobson, S. L., Xi, Z., Rasgon, J. L., Calvitti, M., Moreira, L. A., et al. (2014). Harnessing mosquito-Wolbachia symbiosis for vector and disease control. *Acta Trop.* 132(Suppl.), S150–S163. doi: 10.1016/j.actatropica.2013.11.004

Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J., and Bordenstein, S. R. (2016). Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol.* 14:e2000225. doi: 10.1371/journal.pbio.2000225

Caballero, N. (2016). *Straight from R to JS: Create Interactive Visualizations from R.* Available online at: https://github.com/nachocab/clickme

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2008). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Carvalho-Leandro, D., Ayres, C. F., Guedes, D. R., Suesdek, L., Melo-Santos, M. A., Oliveira, C. F., et al. (2012). Immune transcript variations among *Aedes aegypti* populations with distinct susceptibility to dengue virus serotype 2. *Acta Trop.* 124, 113–119. doi: 10.1016/j.actatropica.2012.07.006

Charan, S. S., Pawar, K. D., Severson, D. W., Patole, M. S., and Shouche, Y. S. (2013). Comparative analysis of midgut bacterial communities of *Aedes aegypti* mosquito strains varying in vector competence to dengue virus. *Parasitol. Res.* 112, 2627–2637. doi: 10.1007/s00436-013-3428-x

Chouaia, B., Rossi, P., Epis, S., Mosca, M., Ricci, I., Damiani, C., et al. (2012). Delayed larval development in Anopheles mosquitoes deprived of Asaia bacterial symbionts. *BMC Microbiol.* 12(Suppl. 1):S2. doi: 10.1186/1471-2180-12-S1-S2

Chouin-Carneiro, T., Vega-Rua, A., Vazeille, M., Yebakima, A., Girod, R., Goindin, D., et al. (2016). Differential susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to zika virus. *PLoS Negl. Trop. Dis.* 10:e0004543. doi: 10.1371/journal.pntd.0004543

Ciota, A. T., Ehrbar, D. J., Matacchiero, A. C., Van Slyke, G. A., and Kramer, L. D. (2013). The evolution of virulence of West Nile virus in a mosquito vector: implications for arbovirus adaptation and evolution. *BMC Evol. Biol.* 13:71. doi: 10.1186/1471-2148-13-71

Ciota, A. T., Matacchiero, A. C., Kilpatrick, A. M., and Kramer, L. D. (2014). The effect of temperature on life history traits of Culex mosquitoes. *J. Med. Entomol.* 51, 55–62. doi: 10.1603/ME13003

Cohen, C., Toh, E., Munro, D., Dong, Q., and Hawlena, H. (2015). Similarities and seasonal variations in bacterial communities from the blood of rodents and from their flea vectors. *ISME J.* 9, 1662–1676. doi: 10.1038/ismej.2014.255

Coon, K. L., Vogel, K. J., Brown, M. R., and Strand, M. R. (2014). Mosquitoes rely on their gut microbiota for development. *Mol. Ecol.* 23, 2727–2739. doi: 10.1111/mec.12771

Correa, C. C., and Ballard, J. W. O. (2016). Wolbachia associations with insects: winning or losing against a master manipulator. *Front. Ecol. Evol.* 3:153. doi: 10.3389/fevo.2015.00153

Crotti, E., Damiani, C., Pajoro, M., Gonella, E., Rizzi, A., Ricci, I., et al. (2009). Asaia, a versatile acetic acid bacterial symbiont, capable of cross-colonizing insects of phylogenetically distant genera and orders. *Environ. Microbiol.* 11, 3252–3264. doi: 10.1111/j.1462-2920.2009.02048.x

Cupp, E. W., Hassan, H. K., Yue, X., Oldland, W. K., Lilley, B. M., and Unnasch, T. R. (2007). West Nile virus infection in mosquitoes in the mid-south USA, 2002-2005. *J. Med. Entomol.* 44, 117–125. doi: 10.1093/jmedent/41.5.117

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., et al. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics* 17:55. doi: 10.1186/s12864-015-2194-9

da Moura, A. J., de Melo Santos, M. A., Oliveira, C. M., Guedes, D. R., de Carvalho-Leandro, D., da Cruz Brito, M. L., et al. (2015). Vector competence of the *Aedes aegypti* population from Santiago Island, Cape Verde, to different serotypes of dengue virus. *Parasit. Vectors* 8, 114. doi: 10.1186/s13071-015-0706-8

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Method* 9, 772. doi: 10.1038/nmeth.2109

Dennison, N. J., Jupatanakul, N., and Dimopoulos, G. (2014). The mosquito microbiota influences vector competence for human pathogens. *Curr. Opin. Insect Sci.* 3, 6–13. doi: 10.1016/j.cois.2014.07.004

Dodson, B. L., Hughes, G. L., Paul, O., Matacchiero, A. C., Kramer, L. D., and Rasgon, J. L. (2014). Wolbachia enhances West Nile virus (WNV) infection in the mosquito Culex tarsalis. *PLoS Negl. Trop. Dis.* 8:e2965. doi: 10.1371/journal.pntd.0002965

Dohm, D. J., O'Guinn, M. L., and Turell, M. J. (2002). Effect of environmental temperature on the ability of *Culex pipiens* (Diptera: Culicidae) to transmit West Nile virus. *J. Med. Entomol.* 39, 221–225. doi: 10.1603/0022-2585-39.1.221

Duguma, D., Hall, M. W., Rugman-Jones, P., Stouthamer, R., Terenius, O., Neufeld, J. D., et al. (2015). Developmental succession of the microbiome of Culex mosquitoes. *BMC Microbiol.* 15:140. doi: 10.1186/s12866-015-0475-8

Dutra, H. L. C., Rocha, M. N., Dias, F. B. S., Mansur, S. B., Caragata, E. P., and Moreira, L. A. (2016). *Wolbachia* blocks currently circulating zika virus isolates in Brazilian *Aedes aegypti* mosquitoes. *Cell Host Microbe* 19, 771–774. doi: 10.1016/j.chom.2016.04.021

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acid Res.* 32, 1792–1297. doi: 10.1093/nar/gkh340

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Faria, V. G., Martins, N. E., Magalhães, S., Paulo, T. F., Nolte, V., Schlötterer, C., et al. (2016). Drosophila adaptation to viral infection through defensive symbiont evolution. *PLoS Genet.* 12:e1006297. doi: 10.1371/journal.pgen.1006297

Floore, T. G. (2006). Mosquito larval control practices: past and present. *J Mosquito Control Assn.* 22, 527–533. doi: 10.2987/8756-971X(2006)22[527:MLCPPA]2.0.CO;2

Gaio, A. de O., Gusmão, D. S., Santos, A. V., Berbert-Molina, M. A., Pimenta, P. F., and Lemos, F. J. (2011). Contribution of midgut bacteria to blood digestion and egg production in *Aedes aegypti* (Diptera: Culicidae) (L.). *Parasit. Vectors* 4:105. doi: 10.1186/1756-3305-4-105

Gimonneau, G., Tchioffo, M. T., Abate, L., Boissière, A., Awono-Ambéné, P. H., Nsango, S. E., et al. (2014). Composition of *Anopheles coluzzii* and *Anopheles gambiae* microbiota from larval to adult stages. *Infect. Genet. Evol.* 28, 715–724. doi: 10.1016/j.meegid.2014.09.029

Glaser, R. L., and Meola, M. A. (2010). The native Wolbachia endosymbionts of *Drosophila melanogaster* and *Culex quinquefasciatus* increase host resistance to West Nile virus infection. *PLoS ONE* 5:e11977. doi: 10.1371/journal.pone.0011977

Guidi, V., Lehner, A., Lüthy, P., and , Tonolla, M. (2013). Dynamics of *Bacillus thuringiensis* var. israelensis and Lysinibacillus sphaericus spores in urban catch basins after simultaneous application against mosquito larvae. *PLoS ONE* 8:e55658. doi: 10.1371/journal.pone.0055658

Hegde, S., Rasgon, J. L., and Hughes, G. L. (2015). The microbiome modulates arbovirus transmission in mosquitoes. *Curr. Opin. Virol.* 15, 97–102. doi: 10.1016/j.coviro.2015.08.011

Hoffmann, A. A., Montgomery, B. L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P. H., Muzzi, F., et al. (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature* 476, 454–457. doi: 10.1038/nature10356

Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, analysis and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046

Hughes, G. L., Dodson, B. L., Johnson, R. M., Murdock, C. C., Tsujimoto, H., Suzuki, Y., et al. (2014). Native microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12498–12503. doi: 10.1073/pnas.1408888111

Hussain, M., Lu, G., Torres, S., Edmonds, J. H., Kay, B. H., Khromykh, A. A., et al. (2013). Effect of Wolbachia on replication of West Nile virus in a mosquito cell line and adult mosquitoes. *J. Virol.* 87, 851–858. doi: 10.1128/JVI.01837-12

Johnson, K. N. (2015). The impact of Wolbachia on virus infection in mosquitoes. *Viruses* 7, 5705–5717. doi: 10.3390/v7112903

Jones, R. T., Knight, R., and Martin, A. P. (2010). Bacterial communities of disease vectors sampled across time, space, and species. *ISME J.* 4, 223–231. doi: 10.1038/ismej.2009.111

Jupatanakul, N., Sim, S., and Dimopoulos, G. (2014). The insect microbiome modulates vector competence for arboviruses. *Viruses* 6, 4294–4313. doi: 10.3390/v6114294

Kambris, Z., Cook, P. E., Phuc, H. K., and Sinkins, S. P. (2009). Immune activation by life-shortening Wolbachia and reduced filarial competence in mosquitoes. *Science* 326, 134–136. doi: 10.1126/science.1177531

Kilpatrick, A. M., Fonseca, D. M., Ebel, G. D., Reddy, M. R., and Kramer, L. D. (2010). Spatial and temporal variation in vector competence of *Culex pipiens* and Cx. restuans mosquitoes for West Nile virus. *Am. J. Trop. Med. Hyg.* 83, 607–613. doi: 10.4269/ajtmh.2010.10-0005

Kueneman, J. G., Parfrey, L. W., Woodhams, D. C., Archer, H. M., Knight, R., and McKenzie, V. J. (2014). The amphibian skin microbiome across species, space and life history stages. *Mol. Ecol.* 23, 1238–1250. doi: 10.1111/mec.12510

Lalzar, I., Harrus, S., Mumcuoglu, K. Y., and Gottlieb, Y. (2012). Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. *Appl. Environ. Microbiol.* 78, 4110–4116. doi: 10.1128/AEM.00323-12

Lanciotti, R. S., Kerst, A. J., Nasci, R. S., Godsey, M. S., Mitchell, C. J., Savage, H. M., et al. (2000). Rapid detection of West Nile virus from human clinical specimens, field collected mosquitoes and avian samples by a TaqMan® RT-PCR assay. *J. Clin. Microbiol.* 38, 4066–4071.

Lu, P., Bian, G., Pan, X., and Xi, Z. (2012). Wolbachia induces density-dependent inhibition to dengue virus in mosquito cells. *PLoS Negl. Trop. Dis.* 6:e1754. doi: 10.1371/journal.pntd.0001754

Martinez, J., Longdon, B., Bauer, S., Chan, Y. S., Miller, W. J., 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. doi: 10.1371/journal.ppat.1004369

Martinez, J., Ok, S., Smith, S., Snoeck, K., Day, J. P., and Jiggins, F. M. (2015). Should symbionts be nice or selfish? Antiviral effects of Wolbachia are costly but reproductive parasitism is not. *PLoS Pathog.* 11:e1005021. doi: 10.1371/journal.ppat.1005021

Mayoral, J. G., Hussain, M., Joubert, D. A., Iturbe-Ormaetxe, I., O'Neill, S. L., and Asgari, S. (2014). Wolbachia small noncoding RNAs and their role in cross-kingdom communications. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18721–18726. doi: 10.1073/pnas.1420131112

Mbewe, R., Pemba, D., Kazembe, L., Mhango, C., and Chiotha, S. (2014). The impact of *Bacillus thuringiensis* israelensis (Bti) on adult and larvae black fly populations. *Malawi J. Sci. Technol.* 10, 86–92.

McMurdie, P. J., and the biom-format team. (2014). *Biom: An Interface Package (Beta) for the BIOM File Format*. Available online at: http://CRAN.R-project.org/package=biom

Micieli, M. V., and Glaser, R. L. (2014). Somatic Wolbachia (Rickettsiales: Rickettsiaceae) levels in *Culex quinquefasciatus* and *Culex pipiens* (Diptera: Culicidae) and resistance to West Nile virus infection. *J. Med. Entomol.* 51, 189–199. doi: 10.1603/ME13152

Minard, G., Mavingui, P., and Moro, C. V. (2013). Diversity and function of bacterial microbiota in the mosquito holobiont. *Parasit. Vectors* 6:146. doi:10.1186/1756-3305-6-146

Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., et al. (2009). A Wolbachia symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and Plasmodium. *Cell* 139, 1268–1278. doi: 10.1016/j.cell.2009.11.042

Murdock, C. C., Moller-Jacobs, L. L., and Thomas, M. B. (2013). Complex environmental drivers of immunity and resistance in malaria mosquitoes. *Proc. R. Soc. B* 280, 20132030. doi: 10.1098/rspb.2013.2030

Muturi, E. J., Kim, C.-H., Bara, J., Bach, E. M., and Siddappaji, M. H. (2016). *Culex pipiens* and *Culex restuans* mosquitoes harbor distinct microbiota dominated by few bacterial taxa. *Parasit. Vectors* 9:18. doi: 10.1186/s13071-016-1299-6

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013). *Vegan: Community Ecology Package*. Available online at: http://CRAN.R-project.org/package=vegan.

Osborne, S. E., Leong, Y. S., O'Neill, S. L., and Johnson, K. N. (2009). Variation in antiviral protection mediated by different Wolbachia strains in *Drosophila simulans*. *PLoS Pathog.* 5:e1000656. doi: 10.1371/journal.ppat.1000656

Osei-Poku, J., Mbogo, C. M., Palmer, W. J., and Jiggins, F. M. (2012). Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Mol. Ecol.* 21, 5138–5150. doi: 10.1111/j.1365-294X.2012.05759.x

Pan, X., Zhou, G., Wu, J., Bian, G., Lu, P., Raikhel, A. S., et al. (2012). Wolbachia induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. U.S.A.* 109, E23–E31. doi: 10.1073/pnas.1116932108

Paris, M., David, J. P., and Despres, L. (2011a). Fitness costs of resistance to Bti toxins in the dengue vector *Aedes aegypti*. *Ecotoxicology* 20, 1184–1194. doi: 10.1007/s10646-011-0663-8

Paris, M., Tetreau, G., Laurent, F., Lelu, M., Despres, L., and David, J. P. (2011b). Persistence of *Bacillus thuringiensis* israelensis (Bti) in the environment induces resistance to multiple Bti toxins in mosquitoes. *Pest Manag. Sci.* 67, 122–128. doi: 10.1002/ps.2046

Primack, R. B. (2014). *Walden Warming*. Chicago, IL: University of Chicago Press.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Reisen, W. K., Fang, Y., and Martinez, V. M. (2006). Effects of temperature on the transmission of west nile virus by *Culex tarsalis* (Diptera: Culicidae). *J. Med. Entomol.* 43, 309–317. doi: 10.1093/jmedent/43.2.309

Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147. doi: 10.1016/0025-5564(81)90043-2

Rossi, P., Ricci, I., Cappelli, A., Damiani, C., Ulissi, U., Mancini, M. V., et al. (2015). Mutual exclusion of Asaia and Wolbachia in the reproductive organs of mosquito vectors. *Parasit. Vectors* 8, 278. doi: 10.1186/s13071-015-0888-0

RStudio Team (2015). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.

Sardelis, M. R., Turell, M. J., Dohm, D. J., and O'Guinn, M. L. (2001). Vector competence of selected North American Culex and Coquilletidia mosquitoes for West Nile virus. *Emer. Infect. Dis.* 7, 1018–1022. doi: 10.3201/eid0706.010617

Sinkins, S. P. (2013). Wolbachia and arbovirus inhibition in mosquitoes. *Future Microbiol.* 8, 1249–1256. doi: 10.2217/fmb.13.95

Stalinski, R., Tetreau, G., Gaude, T., and Després, L. (2014). Pre-selecting resistance against individual Bti Cry toxins facilitates the development of resistanceto the Bti toxins cocktail. *J. Invertebr. Pathol.* 119, 50–53. doi: 10.1016/j.jip.2014.04.002

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sunish, I. P., Rajendran, R., Paramasivan, R., Dhananjeyan, K. J., and Tyagi, B. K. (2011). Wolbachia endobacteria in a natural population of *Culex quinquefasciatus* from filariasis endemic villages of south India and its phylogenetic implication. *Trop. Biomed.* 28, 569–576.

Tchioffo, M. T., Boissière, A., Abate, L., Nsango, S. E., Bayibéki, A. N., Awono-Ambéné, P. H., et al. (2016). Dynamics of bacterial community composition in the malaria mosquito's epithelia. *Front. Microbiol.* 6:1500. doi: 10.3389/fmicb.2015.01500

Van Treuren, W., Ponnusamy, L., Brinkerhoff, R. J., Gonzalez, A., Parobek, C. M., Juliano, J. J., et al. (2015). Variation in the microbiota of Ixodes ticks with regard to geography, species, and sex. *Appl. Environ. Microbiol.* 81, 6200–6209. doi: 10.1128/AEM.01562-15

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Weiss, S. J., Xu, Z., Amir, A., Peddada, S., Bittinger, K., Gonzalez, A., et al. (2015). Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. *PeerJ PrePrints* 3:e1408. doi: 10.7287/peerj.preprints.1157v1

Werren, J. H., Baldo, L., and Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751. doi: 10.1038/nrmicro1969

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer-Verlag. Available online at: http://ggplot2.org doi: 10.1007/978-0-387-98141-3

Wiwatanaratanabutr, I., and Kittayapong, P. (2009). Effects of crowding and temperature on Wolbachia infection density among life cycle stages of *Aedes albopictus. J. Invertebr. Pathol.* 102, 220–224. doi: 10.1016/j.jip.2009.08.009

Yadav, K. K., Bora, A., Datta, S., Chandel, K., Gogoi, H. K., Prasad, G. B., et al. (2015). Molecular characterization of midgut microbiota of *Aedes albopictus* and *Aedes aegypti* from Arunachal Pradesh, India. *Parasit. Vectors* 8:641. doi: 10.1186/s13071-015-1252-0

Ye, Y. H., Carrasco, A. M., Dong, Y., Sgrò, C. M., and McGraw, E. A. (2016). The effect of temperature on Wolbachia-mediated Dengue virus blocking in *Aedes aegypti*. *Am. J. Trop. Med. Hyg.* 94, 812–819. doi: 10.4269/ajtmh.15-0801

Zélé, F., Nicot, A., Berthomieu, A., Weill, M., Duron, O, and ., Rivero, A. (2014). Wolbachia increases susceptibility to Plasmodium infection in a natural system. *Proc. R. Soc. B* 281, 20132837. doi: 10.1098/rspb.2013.2837

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Paper 12:

Rodriguez-Ruano, S. M., Škochová, V., Rego, R. O. M., Schmidt, J. O., Roachell, W., Hypša, V., & **Novakova, E.** (2018). Microbiomes of North American Triatominae: the grounds for Chagas disease epidemiology. *Frontiers in Microbiology*, *9*, 1167.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 June 2018
doi: 10.3389/fmicb.2018.01167

Microbiomes of North American Triatominae: The Grounds for Chagas Disease Epidemiology

Sonia M. Rodríguez-Ruano[1], *Veronika Škochová*[1], *Ryan O. M. Rego*[2], *Justin O. Schmidt*[3], *Walter Roachell*[4], *Václav Hypša*[1,2] *and Eva Nováková*[1,2]*

[1] Faculty of Science, University of South Bohemia, Ceske Budejovice, Czechia, [2] Biology Centre of ASCR, Institute of Parasitology, Ceske Budejovice, Czechia, [3] Department of Entomology, Southwestern Biological Institute, Tucson, AZ, United States, [4] US Army Public Health Command-Central, JBSA Fort Sam Houston, Houston, TX, United States

Insect microbiomes influence many fundamental host traits, including functions of practical significance such as their capacity as vectors to transmit parasites and pathogens. The knowledge on the diversity and development of the gut microbiomes in various blood feeding insects is thus crucial not only for theoretical purposes, but also for the development of better disease control strategies. In Triatominae (Heteroptera: Reduviidae), the blood feeding vectors of Chagas disease in South America and parts of North America, the investigation of the microbiomes is in its infancy. The few studies done on microbiomes of South American Triatominae species indicate a relatively low taxonomic diversity and a high host specificity. We designed a comparative survey to serve several purposes: (I) to obtain a better insight into the overall microbiome diversity in different species, (II) to check the long term stability of the interspecific differences, (III) to describe the ontogenetic changes of the microbiome, and (IV) to determine the potential correlation between microbiome composition and presence of *Trypanosoma cruzi*, the causative agent of Chagas disease. Using 16S amplicons of two abundant species from the southern US, and four laboratory reared colonies, we showed that the microbiome composition is determined by host species, rather than locality or environment. The OTUs (Operational Taxonomic Units) determination confirms a low microbiome diversity, with 12-17 main OTUs detected in wild populations of *T. sanguisuga* and *T. protracta*. Among the dominant bacterial taxa are *Acinetobacter* and *Proteiniphilum* but also the symbiotic bacterium *Arsenophonus triatominarum*, previously believed to only live intracellularly. The possibility of ontogenetic microbiome changes was evaluated in all six developmental stages and feces of the laboratory reared model *Rhodnius prolixus*. We detected considerable changes along the host's ontogeny, including clear trends in the abundance variation of the three dominant bacteria, namely *Enterococcus*, *Acinetobacter,* and *Arsenophonus*. Finally, we screened the samples for the presence of *Trypanosoma cruzi*. Comparing the parasite presence with the microbiome composition, we assessed the possible significance of the latter in the epidemiology of the disease. Particularly, we found a trend toward more diverse microbiomes in *Trypanosoma cruzi* positive *T. protracta* specimens.

Keywords: Triatominae, microbiome, ontogeny, *Trypanosoma cruzi*, *Rhodnius prolixus*

OPEN ACCESS

Edited by:
Francesca Turroni,
Università degli Studi di Parma, Italy

Reviewed by:
Guillaume Henri Tetreau,
UMR5075 Institut de Biologie Structurale (IBS), France
Alex Cordoba-Aguilar,
Universidad Nacional Autónoma de México, Mexico

***Correspondence:**
Eva Nováková
novaeva@paru.cas.cz

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *09 March 2018*
Accepted: *14 May 2018*
Published: *13 June 2018*

Citation:
Rodríguez-Ruano SM, Škochová V, Rego ROM, Schmidt JO, Roachell W, Hypša V and Nováková E (2018) Microbiomes of North American Triatominae: The Grounds for Chagas Disease Epidemiology. Front. Microbiol. 9:1167. doi: 10.3389/fmicb.2018.01167

INTRODUCTION

Insect microbiomes are recognized as a major factor determining host fitness and various phenotypic traits, including vectorial capacity in blood feeding species (Klepzig et al., 2009; Minard et al., 2013; Oliver and Martinez, 2014). While vector control strategies are now naturally venturing into the microbiome concept (Crotti et al., 2012; Saldaña et al., 2017), for some insect vectors and the allied pathogens the research focus has not yet fully passed the borders of epidemiological, surveillance, and medical studies. One such example are kissing bugs of the subfamily Triatominae (Hemiptera: Reduviidae) and *Trypanosoma cruzi*, the protozoan parasite responsible for Chagas disease.

Triatominae are hemimetabolous insects with five nymphal instars. Both adults and nymphs feed on vertebrate blood. There are over 140 extant Triatominae species with different biology, vectorial capacity and distribution, most of them found in South and Central America (Galvão et al., 2003; de Fuentes-Vicente et al., 2018). Those areas are the hot spots for Chagas disease, with nearly six million people affected (WHO, 2015). Among the seven main vector species of *Trypanosoma cruzi* in Latin America (Hernández et al., 2016), the most attention has traditionally been paid to *Rhodnius prolixus* colonizing domestic habitats (Gourbière et al., 2012). For many decades, *R. prolixus* has also served as a model organism to study insect physiology, immunity, metabolism, and development (Nunes-da-Fonseca et al., 2017). Such a status has promoted a high availability of information on this Triatominae species, including a whole genome sequence (Mesquita et al., 2015), and has contributed to unraveling its role as the vector for *Trypanosoma cruzi*.

The southern part of the United States is also endemic for this trypanosomiasis (Montgomery et al., 2014) and Chagas disease has recently been listed among the emerging diseases in this region (Edwards et al., 2017). Eleven Triatominae species are considered native to North America (Bern et al., 2011). The five species most frequently found are *Triatoma gerstaeckeri*, *T. indictiva, T. lecticularia, T. sanguisuga, and T. rubida* (Curtis-Robles et al., 2017a). So far, a modest number of both human and canine *Trypanosoma cruzi* infections has been documented in Texas (Wozniak et al., 2015; Curtis-Robles et al., 2017a). The latest surveillance data, however, reveal the presence of Triatominae adults and nymphs in domestic and peridomestic areas, which points at a potential increase of human infections (Curtis-Robles et al., 2017a). In fact, current studies are already describing a growing number of human subjects locally infected with *Trypanosoma cruzi* in the southern US (Garcia et al., 2017; Gunter et al., 2017).

Since *Trypanosoma cruzi* undergoes development exclusively in the intestinal tract of Triatominae vectors, it is exposed to interactions with gut microbes. Many studies to date have revealed an effect (either positive or negative) of microbiomes on the vectorial capacity of their insect hosts through different mechanisms such as resource competition, immunity modulation or antiparasitic secondary metabolite production (Cirimotich et al., 2011; Jupatanakul et al., 2014; Hegde et al., 2015). In Triatominae, the microbiomes interact with *Trypanosoma cruzi* by direct contact, including competition for resources (Garcia et al., 2010), or indirectly by increasing the expression of antiparasitic molecules and humoral immune defense factors of the host (Azambuja et al., 2005; Weiss and Aksoy, 2011). While there are two major threads to commonly used chemical control of Triatominae vectors, i.e., developing resistance and recurrent infestations by sylvatic species, symbiont manipulation has been proposed as a sustainable alternative (Gourbière et al., 2012). For instance, trypanolytic activity of some Triatominae associated *Serratia marcescens* strains (Azambuja et al., 2004) and gene silencing using transformed *Rhodococcus rhodnii* have been explored (Taracena et al., 2015). However, any advancement in microbiome-based vector control is hampered by the lack of a solid background on Triatominae symbiosis arising from the basic knowledge on the microbiome characteristics.

Until now, microbiome data and their basic characteristics have been available for a limited number of South American Triatominae species from wild and laboratory colonies (da Mota et al., 2012; Gumiel et al., 2015; Díaz et al., 2016). Based on these data, Triatominae possess low-complexity microbiomes, with one or few dominant bacterial genera, which appear to be specific to certain hosts (da Mota et al., 2012; Díaz et al., 2016). The complexity of lab-reared insect microbiomes is even lower, but maintains most of the bacterial groups found in their wild counterparts (da Mota et al., 2012). While for other insects the microbiomes are known to vary along the host ontogeny (Sudakaran et al., 2012; Gimonneau et al., 2014; Duguma et al., 2015), the development of Triatominae microbiomes is completely unknown. Considering the ability of all the instars to transmit *Trypanosoma cruzi* (Wozniak et al., 2015), possible changes in the microbiome composition along their ontogeny may be a key point toward the understanding of the disease transmission and its biological control. In this study, we extend the analysis of microbiome diversity and specificity to North American species from wild and laboratory colonies. We also take advantage of the model species *R. prolixus* and describe the microbiome variation in its different developmental stages. Finally, we present the first data on the correlation between the microbiome richness and presence of *Trypanosoma cruzi* in wild populations of two broadly found North American species, *T. protracta* and *T. sanguisuga*.

MATERIALS AND METHODS

The Design of the Sample Set: Taxonomy and Origin

We used samples of four Triatominae species reared in long term laboratory colonies (*R. prolixus*, *T. vitticeps*, *T. protracta,* and *T. recurva*) and field collected samples of two species, *T. protracta* and *T. sanguisuga*. Due to this arrangement we could address the main hypothesis (i.e., the host specific microbiome composition) in two different ways. First, we compared field sampled individuals of two species from different localities to see whether the microbiome composition is determined by the host taxonomy or rather the geography. Second, we compared microbiomes of

four species from laboratory colonies to determine whether the significant interspecific differences can be detected after a long term cohabitation in the same environment. Finally, since it has been shown in some other arthropods (the phylogenetically closest example being heteropterans of the genus Pyrrhocoris, see Discussion) that the microbiome composition can change during the host ontogeny, we used an available *R. prolixus* colony to directly characterize the ontogenetic development of the microbial communities across different developmental stages in a closely related and ecologically similar model.

Rearing Conditions and Microbiome Ontogeny

The microbiome ontogeny was assessed as follows. Eggs, all five nymphal instars (L1-L5), and adults of *R. prolixus* were analyzed in triplicates as described hereafter. Upon feces availability, feces from adults were collected on a filter paper within 24 h after deposition and processed in five replicates. The *R. prolixus* colony, established in 2010, originates from Brazil and has since been kept at the Faculty of Science, University of South Bohemia, Czech Republic. The colony is reared in plastic containers with mesh lids at a constant temperature of 27°C and 70% humidity under 16:8 h light/dark cycles. The blood meal, commercially purchased defibrinated sheep blood supplemented with egg yolk (Núñez and Lazzari, 1990), is supplied in 10 day intervals using artificial membrane feeding *ad libitum*.

Interspecific Diversity and Stability of the Microbiomes

The variability and host specificity of the microbial communities (i.e., the assumption that the specific microbiome profile is determined by the host species), was tested using laboratory reared colonies. In order to prevent downsizing of the colonies, only first and second instars were sampled in five replicates from four Triatominae species. The origins of the samples were as follows (see **Table 1**). The *R. prolixus* colony and *T. vitticeps* individuals were acquired from the above-described facility and were kept under the same breeding conditions. The colony of *T. vitticeps* originates from field-captured individuals sampled in Brazil and kept since 1995. The two other species, *T. protracta* and *T. recurva,* representing the North American (NA) Triatominae, were provided by the Southwestern Biological Institute, Tucson, AZ, USA. All the NA colonies were established by Justin Schmidt in 2010. *T. protracta* originated from Tucson, AZ, while *T. recurva* was sampled in Bisbee, AZ. The colonies are kept under simulated natural environmental conditions experienced by the insects in their native habitat. The temperature in the insectary is allowed to fluctuate with the weather, except with heating to maintain at least 4°C and cooling to maintain a temperature no higher than 32°C. Humidity is not being modified except during cooling conditions when maintained at 30%. Lighting is natural as provided by numerous windows of the insectary. Meals of rodent blood are provided every 14 days *ad libitum* as approved by the institutional review board of the Southwestern Biological Institute.

In addition to the laboratory-reared samples, we analyzed field collected samples to gain an initial insight into the diversity and ecology of the microbiome associated with the most widespread NA species. Triatominae sampled in the three most affected states, i.e., California, Arizona, and Texas, were provided by Christiane Weirauch (UC Riverside, California) and Walter D. Roachell (US Army Public Health Command Central, Texas). The samples were morphologically assigned to two species: *T. sanguisuga* and *T. protracta*, and the determination was further confirmed by molecular data. The complete information on laboratory bred and field collected individuals used in this study is provided in **Table 1**.

Template Extraction, Molecular Taxonomy and *Trypanosoma cruzi* Diagnostics

Colony individuals were sampled 4 days after feeding. Due to their size, the first instars were not dissected, and whole surface sterilized abdomens were used (see below for sterilization details). For the second and older instars in the ontogenetic survey, gut samples were acquired using a simple dissection protocol. The insects were surface sterilized by submersion in absolute ethanol for 5 s followed by a rinse in sterile deionizied water and dissected in sterile phosphate buffer saline (1X PBS; Sigma Life Science). The ventral cuticle was removed and the entire length of the gut was collected into individual tubes for each DNA extraction. The dissection ensured acquiring bacteria predominating the gut, where the main interactions with the host and possible parasites are expected to occur, and it also reduced any external contamination. To reduce possible contamination of the eggs, their environmental exposure was minimized collecting them immediately after deposition directly into tubes with the extraction buffer.

For the field sampled individuals, we optimized the extraction protocol so that comparable microbiome profiles could be generated from ethanol preserved abdomens (provided by some of our collaborators and available in numerous collections worldwide) and fresh dissected samples. Two separate templates were used from the dissected samples, i.e., the gut and the other abdominal tissues and organs (**Table 1**). This protocol, followed by merging the corresponding microbiome data into a "bioinformatic abdomen" (see section Data Processing and Statistical Analyses), ensures data comparability and provides gDNA gut specific templates for any future applications (i.e., qPCR and metagenomics). Total DNA was extracted from all the above described templates (**Table 1**) using the DNeasy Blood and Tissue (Qiagen, Hilden, Germany).

In order to verify the taxonomical determination of the 29 field collected individuals, we sequenced fragments of the COI gene amplified with the forward primers COIL6625 and/or COIL1490 and the reverse primer COIH7005 (Hafner et al., 1994). The reference sequence for *T. sanguisuga* was obtained from the adult individuals sampled by Walter Roachel in Lackland Air Force Base area, TX. The reference for *T. protracta* was sequenced from the laboratory colony (see above). A Maximum Likelihood (ML) analysis was performed with PhyML

TABLE 1 | List of all the DNA isolates for which the 16S rRNA gene amplicon data were generated in this study.

Species	Instar, sex	No. of samples	DNA template	Origin
R. prolixus	L1	3	Whole abdomen	Laboratory colony, CZ
	L2	3	Gut	Laboratory colony, CZ
	L3	3	Gut	Laboratory colony, CZ
	L4	3	Gut	Laboratory colony, CZ
	L5	3	Gut	Laboratory colony, CZ
	adult, M	1	Gut	Laboratory colony, CZ
	adult, F	2	Gut	Laboratory colony, CZ
	NA	3	Egg	Laboratory colony, CZ
	NA	5	feces	Laboratory colony, CZ
T. vitticeps	L1	3	Whole abdomen	Laboratory colony, CZ
	L2	3	Gut	Laboratory colony, CZ
T. protracta	L1	5	Whole abdomen	Laboratory colony, USA
T. recurva	L1	5	Whole abdomen	Laboratory colony, USA
T. sanguisuga	L4	4	Gut	San Antonio, TX, 2016
	L5	3	Gut	San Antonio, TX, 2016
	adult, F	2	Gut	San Antonio, TX, 2016
	adult, M	1	Gut	San Antonio, TX, 2016
	L4	4	Other tissues and organs	San Antonio, TX, 2016
	L5	3	Other tissues and organs	San Antonio, TX, 2016
	adult, F	2	Other tissues and organs	San Antonio, TX, 2016
	adult, M	1	Other tissues and organs	San Antonio, TX, 2016
T. protracta	adult, M	7	Whole abdomen	Escondido, CA, 2007
	adult, F	7	Whole abdomen	Escondido, CA, 2007
	adult, M	2	Whole abdomen	Tucson, AZ, 2016
	adult, F	1	Whole abdomen	Tucson, AZ, 2016
	adult, M	1	Other tissues and organs	San Antonio, TX, 2016
	adult, M	1	Gut	San Antonio, TX, 2016
	adult, F	1	Other tissues and organs	San Antonio, TX, 2016
	adult, F	1	Gut	San Antonio, TX, 2016

F and M stands for female and male. The bioinformatic abdomens were generated merging the Gut data and the data from Other tissues and organs for every respective individual.

plugin in Geneious (Guindon et al., 2010; Kearse et al., 2012) to confirm the taxonomical assignment.

The field collected individuals were further screened for the presence of *Trypanosoma cruzi*. The diversity of *Trypanosoma cruzi* discrete typing units (DTUs), associated with geography, vector species, transmission and clinical outcomes (Hernández et al., 2016), spans five out of the six designated DTUs within United States (Garcia et al., 2017). Three sets of diagnostic primers were thus employed in this study: one generally amplifying all *Trypanosoma cruzi* DTUs (TCZ1/TCZ2; Moser et al., 1989) and the two other distinguishing for DTUI and the DTUII-VI group (ME/TC1 or TC2; Gumiel et al., 2015).

Library Preparation and Microbiome Sequencing

To generate data suitable for microbiome profiling, gDNA of 83 samples was amplified according to the EMP protocol (http://www.earthmicrobiome.org/protocols-and-standards/16s/), producing approximately 250 bp of the 16S rDNA V4 hypervariable region. Two negative controls of the extraction procedure and one blank control for PCR amplification were included in the sequencing. The libraries were then sequenced in a single run of Illumina MiSeq using the v2 chemistry (2 × 150 bp output mode).

Data Processing and Statistical Analyses

Forward and reverse paired-end reads were merged using USEARCH v7.0.1001 (fastq_mergepairs with fastq_minovlen set to 20; Edgar, 2013). Then the sequences were demultiplexed and their quality was checked using QIIME 1.8 (split_libraries_fastq.py with phred_quality_threshold set to 19; Caporaso et al., 2010b). The retained high quality sequences were aligned using the QIIME implementation of Pynast (Caporaso et al., 2010a) and trimmed to a final length of 251 bp using USEARCH (Edgar, 2013). Finally, the dataset was clustered at 100% identity to get a representative set for *de novo* OTU picking using the USEARCH global alignment option at 97% identity (Edgar, 2013). Each OTU was assigned to different taxonomic levels using the BLAST algorithm (Camacho et al., 2009) against the SILVA 123 database (Quast et al., 2013). The obtained OTU table was filtered to remove singletons

and very low abundant OTUs using QIIME (following the recommendations of Bokulich et al., 2013), and all non-bacterial, chloroplast and mitochondrial OTUs were filtered out. After this final filtering step, the OTU table was rarefied at 1000 sequences per sample to normalize the dataset (as recommended in Weiss et al., 2017). The unrarefied data obtained for guts and other internal organs/tissues from each of the dissected wild sampled individuals were merged into "bioinformatic abdomens" using the QIIME script *collapse_samples.py* and normalized at the same rarefaction level of 1000 sequences per sample.

All statistical tests were performed in R environment (R Development Core Team, 2014), mainly using functions from the vegan package (Oksanen et al., 2013). The overall similarity among microbiomes was analyzed based on Bray-Curtis dissimilarities and visualized in two dimensional space using Non-Metric Dimensional Scaling (NMDS; Minchin, 1987). Using the vegan Adonis function, we tested the statistical significance of the differences among microbiomes from distinct host instars and host species. For *T. protracta*, differences between the distribution of diversity indices (Shannon index and richness) of males and females and between *Trypanosoma cruzi* positive and negative individuals were statistically evaluated using Kruskal-Wallis rank test (Kruskal and Wallis, 1952). All graphical outputs were generated using ggplot2 package (Wickham, 2009). In addition, 95% confident ellipses (Fox and Weisberg, 2011) were calculated with the *stat_ellipse* function of the same package.

RESULTS

Sanger and Illumina Data

For molecular determination, we obtained 25 fragments 381 bp long and 4 fragments 652 bp long of the COI sequences. ML analysis of these fragments (not shown) determined 10 individuals as *T. sanguisuga* and 19 as *T. protracta*. The representative sequences for each species were deposited in GenBank under the following accession numbers: MH025901 and MH025902.

Illumina data were obtained for 42 samples from the laboratory reared colonies and 29 wild sampled individuals and 3 negative controls. The mean number of high quality merged reads passing the QC was 9279 ($SD = 554$). The extraction procedure controls resulted in low read numbers (988 and 666). The first extraction control was dominated by OTU161 (*Segetibacter*) that did not occur in any of the Triatominae microbiome data. The second extraction control sample as well as the blank PCR sample (253 reads) were composed of a mixture of low abundant OTUs that were excluded from further analysis. The raw data are available under ENA project no. PRJEB25175.

Microbiome Development Throughout the Host Ontogeny

The microbiome of *R. prolixus* undergoes major compositional changes along the host ontogeny (**Figures 1**, **2**). Significant differences were found when comparing the microbiome of eggs, first four instars, fifth instar and adults, and feces of this species ($R^2 = 0.61335$, significance level of 0.001, **Figure 1B**). When compared pairwise, all the microbiome pairs bore significant differences (four pairs at 0.01 level of significance and two at 0.05, see Supplementary Table 1). On average, the egg replicates showed the highest richness: 37 OTUs represented by numerous aerobes and facultative anaerobes. Two taxa equally dominated the microbial composition, namely *Enterococcus* and *Bartonella* (each accounting for 21.8%). The third well represented taxon was *Methylobacterium* (18.2%).

Starting with the first nymphal instar the composition of the bacterial community changes dramatically (**Figure 2**). The richness of the microbiome drops in the nymphal stages to 13-19 OTUs (**Figure 2B**), and in the feces it reaches 92 OTUs. During the ontogeny from L1 to an adult, the composition of the microbiome shows several consistent trends. The most pronounced of these trends are expressed in the quantities of the three dominant OTUs, *Enterococcus*, *Acinetobacter,* and *Arsenophonus* (**Figure 2B**). While the proportion of *Enterococcus*/*Acinetobacter* in the microbiome steadily increases/decreases during the host ontogeny, the number of *Arsenophonus* reaches its peaks in L3 and then starts decreasing (**Figure 2**). The only dominant OTU without clear consistent pattern was *Macrococcus* which varied in numbers among the nymphal stages and was absent in adults.

Microbiomes of *T. protracta* and *T. sanguisuga*

The two species from wild populations showed major differences between their microbiome profiles at the significance level of 0.001 ($R^2 = 0.289$, **Figure 3**). Since the samples of *T. protracta* were mostly alcohol preserved, the two species were compared based on the microbial data retrieved from the entire abdomens and bioinformatically merged abdomen data. *T. sanguisuga* microbiomes retrieved from individuals of late nymphs (L4 and L5) and adults were composed of 17 OTUs with an averaged relative abundance higher than 1% (Supplementary Table 2). Among these, *Arsenophonus* and *Acinetobacter* bacteria clearly dominated. On average the relative abundance of *Arsenophonus* in all analyzed *T. sanguisuga* individuals was 39.5%, reaching 82.5% in the L4 nymphs, while for *Acinetobacter* the average was 14.3% with the highest relative abundance value of 34.1% in the last nymphal instar L5. Four other taxa exhibited an average relative abundance over 3%, i.e., *Stenotrophomonas* (6.2%), *Rhodopseudomonas* (4.6%), *Pseudomonas* (4.4%), and *Sphingobacterium* (3.4%) (Supplementary Table 2). *T. protracta* microbiome of 16 adult individuals was composed of 12 OTUs with an average relative abundance higher than 1% and the dominating OTU assigned to the genus *Proteiniphilum* (40.7%). Five OTUs showed an average relative abundance above 3%, i.e., *Dietzia* (11.6%), *Salmonella* (10.4%), *Peptoniphilus* (9.0%), Neisseriaceae (6.1%), *Mycobacterium* (3.7%). *Dietzia* was however recovered from a single individual of *T. protracta* reaching the abundance of 97% of all reads generated from this particular specimen (removed as an outlier from the *Trypanosoma cruzi* related analysis).

When testing possible variability of *T. protracta* microbiome composition in relation to the sex of the host individuals,

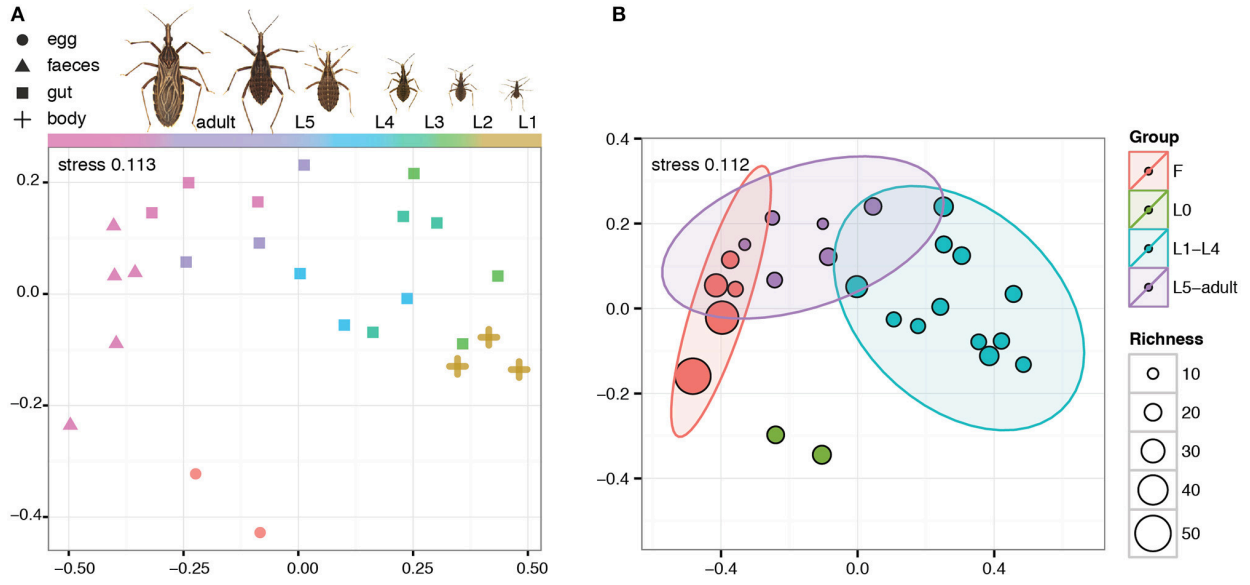


FIGURE 1 | NMDS visualization of the ontogenetic shift of *R. prolixus* microbiome. F stands for feces, L0 for eggs, L1–L5 for nymphal instars. **(A)** Microbiome shift along the different developmental stages of *R. prolixus*. **(B)** NMDS including microbiome richness and 95% confident ellipses for each developmental stage. The ellipses were calculated with the *stat_ellipse* function of ggplot2 package.

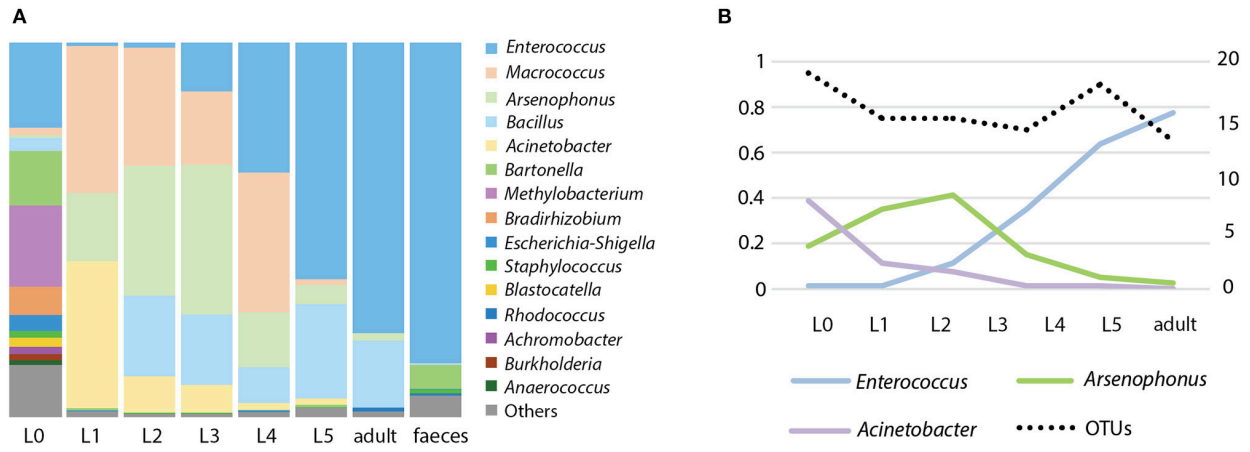


FIGURE 2 | Microbiome of *R. prolixus* along its ontogeny. **(A)** Relative abundance of the main genera in different developmental stages. L0 stands for eggs, L1-L5 are nymphal instars. **(B)** Variation of the dominant OTUs; percentage of the total reads number (left axis) and number of OTUs in individual stages (right axis).

their geographical origin or the year of sampling, no significant differences were found. The only significant correlation was identified for the presence of *Trypanosoma cruzi* and the richness of *T. protracta* microbiome (Kruskal-Wallis chi-squared = 7.4558, df = 1, *p*-value = 0.006; **Figure 4**). The prevalence of *Trypanosoma cruzi* in 16 analyzed *T. protracta* adult individuals was 25%, with two cases of DTUI and the other two from DTUII-VI group.

Due to a single case of *Trypanosoma cruzi* infection in the tested set of *T. sanguisuga*, any correlation of microbiome profile with the presence of the parasite could not be evaluated for this species. It is however noteworthy that this case represented a double infection, with positive diagnostic PCR for both DTUI and DTUII-VI groups.

Microbiomes in the Laboratory Reared Colonies

The laboratory reared colonies also showed differential patterns in their microbiomes according to the Triatominae host species. Particularly, differences among the four species were evaluated as significant at the level of 0.001 (R^2 = 0.784) using adonis (**Figure 5**). Colonies of *T. protracta* and *T. recurva* were dominated by Enterobacteriaceae (*Pectobacterium* sp.), while colonies of *T. vitticeps* were

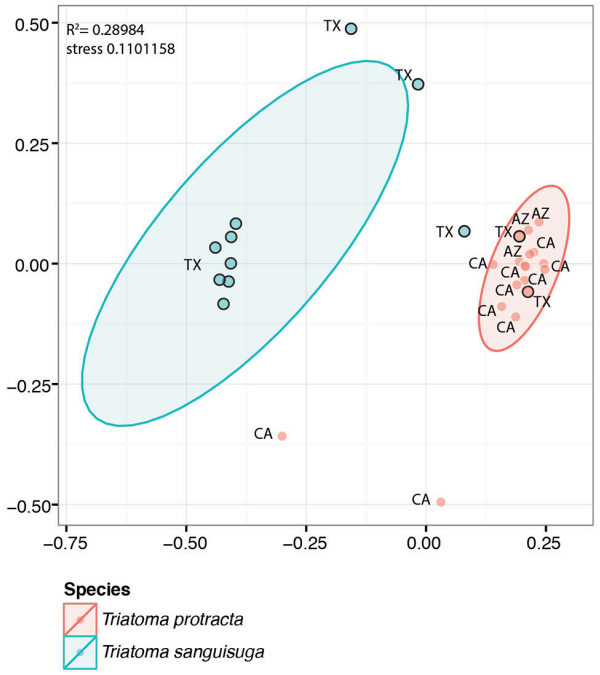


FIGURE 3 | NMDS visualization of significant differences between the microbiomes of *T. protracta* and *T. sanguisuga* sampled from wild populations. The abbreviations stand for the states where the bugs were sampled. The black border circles designate samples for which the bioinformatic abdomen was generated (see Material and Methods section). The ellipses are 95% confident ellipses calculated with the *stat_ellipse* function of ggplot2 package.

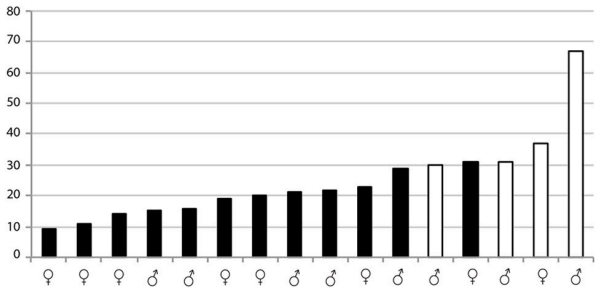


FIGURE 4 | Significant differences in microbiome richness of *Trypanosoma cruzi* infected and uninfected *T. protracta* adult individuals measured by Kruskal-Wallis rank test (chi-squared = 7.4558, df = 1, *p*-value = 0.006323). The numbers stand for the richness values. The black bars represent *Trypanosoma cruzi* negative samples. The white bars are *Trypanosoma cruzi* positive samples.

dominated mainly by Enterococcaceae (*Enterococcus* sp.) and also by Enterobacteriaceae (*Arsenophonus* sp.). *R. prolixus* did not show one single predominant OTU, but three with similar high abundance: Staphylococcaceae (*Macrococcus* sp.), Enterobacteriaceae (*Arsenophonus* sp.) and Moraxellaceae (*Acinetobacter* sp.). Bacillaceae (*Bacillus* sp.) were the next more predominant bacteria in *R. prolixus* and *T. vitticeps*.

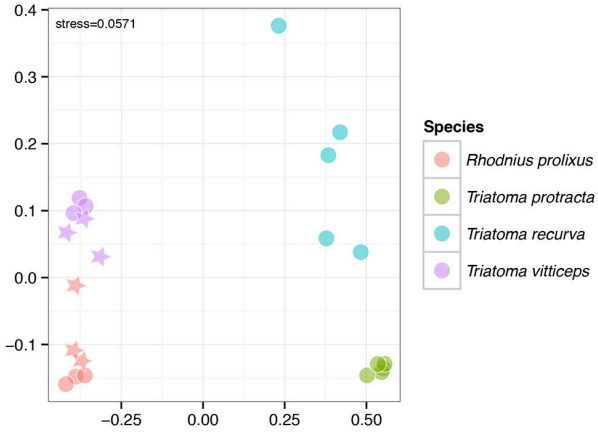


FIGURE 5 | Significant differences among microbiomes of the lab-reared colonies depicted by the NMDS clustering (significance level of 0.001, R^2 = 0.784). The circles stand for L1 instar, stars for L2.

Arsenophonus triatominarum Patterns

Both field collected and laboratory reared Triatominae showed distinct patterns for *A. triatominarum* presence. The averaged relative abundance of this bacterium in the gut of *T. sanguisuga* field-collected individuals was 39.76% (*SD* = 40.19%). However, the relative abundance showed an immense variance among the analyzed individuals spanning from as low as 0.02% found in L5 instars up to 99.03% in the adults. No reads for *A. triatominarum* were detected in *T. protracta* from wild populations. For laboratory-reared colonies, the patterns of *A. triatominarum* were more variable: *R. prolixus* harbored *A. triatominarum* in the gut (26.36 ± 20.43%), but this bacterium was absent in the samples from eggs or feces. *T. vitticeps* also harbored *Arsenophonus* in the gut (29.54 ± 28.21%). *Arsenophonus* was not found in *T. recurva* and *T. protracta* microbiomes of the sampled colonies.

DISCUSSION

Microbiomes and Host Ontogeny

Given the current knowledge, it is difficult to make any general comparison of the pattern we found in Triatominae (i.e., the microbiome of early nymphs being different from the adults one) to other arthropods, since only few other groups have been analyzed to date. For example in ticks, the microbiome changes along the development toward higher diversity, probably due to off-host environmental bacterial acquisition (Zolnik et al., 2016). However, ticks feed on blood throughout their life, and the main component of the microbiome, *Rickettsia*, is present in all life stages, particularly in adult females and larvae, where it comprises most of the microbiome, possibly due to vertical transmission mechanisms (Zolnik et al., 2016). A different ontogenetic pattern was detected in mosquitoes, whose larval microbiome is more diverse and highly influenced by the environment (i.e., aquatic), while the adult microbiome shows a different, more stable pattern (Gimonneau et al., 2014;

Duguma et al., 2015). As mosquitoes are hematophagous only in the adult stage, it has been suggested that selection toward functionally important bacteria in this context acts to shape the adult female mosquito microbiome (Gimonneau et al., 2014). Within Heteroptera, the bloodfeeding Cimicidae do not rely on gut microbiome, but possess so called "primary symbionts" dwelling in bacteriomes (Hypša and Aksoy, 2003; Nikoh et al., 2014), which make them difficult to compare to Triatominae. Perhaps the most similar pattern has been found in non-hematophagous heteropterans, the firebug *Pyrrhocoris apterus*. This species also show remarkable shifts in the microbiome profile when comparing the first instars with the latter instars and adults, but this difference is probably due to the change of feeding habits between stages (Sudakaran et al., 2012). There is currently too little data to make any conclusion on the phylogeny or feeding strategy as the main determinant of the microbiome ontogeny, but it is interesting to note that even in strictly hematophagous vectors, the blood meal effect seems to be quite limited (Zolnik et al., 2016).

Regarding the higher richness of bacteria in feces and eggs, the environmental component is worth discussion. The higher diversity in the egg microbiome compared to the nymphal instars could reflect a common contact of eggs and feces. In addition, the feces samples were exposed to possible external contamination and aerobic bacterial growth for a longer period (up to 24 h compared to the immediate collection of eggs after deposition). The three-fold increment in the richness of feces microbiome could thus be explained by this exposure time difference.

Triatominae can act as vectors of *Trypanosoma cruzi* at any of their developmental stages. The prevalence of *Trypanosoma cruzi* infection seems to be higher in adults and the latest instars compared to the younger nymphs, although this is based on a small number of unidentified nymph samples (McPhatter et al., 2012; Curtis-Robles et al., 2015; Wozniak et al., 2015). Different exposure to blood feeding on different hosts might be one of the possible factors behind these different infection rates. In our analyses of laboratory-reared individuals, we identify microbiome composition as another important trait which undergoes significant changes during the vector's ontogeny. This suggests that the microbiome may be another possible determinant of *Trypanosoma cruzi* prevalence. A further investigation on the microbiome ontogenetic patterns in wild populations of Triatominae and its correlation with *Trypanosoma cruzi* infection is needed to confirm this hypothesis.

Microbiomes in Wild Populations

Our results support previous findings on low-level complexity microbiomes of South American Triatominae. The most abundant bacterial taxa in *T. sanguisuga* and *T. protracta* comply with the diversity span previously described for South American species. *T. sanguisuga* microbiome profile dominated by two taxa, i.e., *Arsenophonus* and *Acinetobacter,* mirrors for instance that of *T. brasiliensis* complex (Díaz et al., 2016). *T. protracta* harbors distinct bacterial lineages. While *Salmonella* association with some Triatominae species was previously reported (Amino et al., 1998), additional taxa found here in *T. protracta* microbiome, i.e., Porphyromonadaceae *(Proteiniphilum* sp.), Clostridiales (*Peptoniphilus* sp.) and Neisseriaceae were identified as symbionts in other insects (e.g., Kwong et al., 2014; Husseneder et al., 2017; Zhang et al., 2017).

Microbiome species-specificity have been detected in several insect hosts to date (Aksoy et al., 2014; Leonhardt and Kaltenpoth, 2014; Novakova et al., 2017). The microbiome of Triatominae does not seem to be an exception, and species-specific patterns have been described as one of the main factors shaping the microbiome composition (da Mota et al., 2012; Díaz et al., 2016). While these studies were conducted mostly using laboratory-reared bugs, our results point out microbiome species-specific patterns in wild populations of NA Triatominae.

Apart from the species-specific patterns, we also checked for other factors possibly affecting the microbiome composition in Triatominae. *T. protracta* was collected from three remote locations in the southern US (San Antonio, TX; Tucson, AZ; Escondido, CA). However, neither the geographical origin of the samples nor the sex of the adult individuals showed an effect on the Triatominae microbiome. These results are in accordance with the characteristics described for another hemipteran, the firebug, where the microbiome is remarkably stable across different populations, sexes, and even diets (Sudakaran et al., 2012). However, since our data lack the information on blood meal source and originate from a complex sample set, including L4, L5 and adult individuals, the effects of these variables on the microbiome composition within wild populations will require further investigations.

Microbiome—*Trypanosoma cruzi* Interface

The prevalence of *Trypanosoma cruzi* in our *T. protracta* samples was 25%, lower than the prevalence found in previous studies of Triatominae from Texas (Curtis-Robles et al., 2017b). However, we still identified different *Trypanosoma cruzi* DTUs (here only differentiated as DTUI and non-DTUI). These findings are in accordance with recent studies, where at least four different DTUs have been found in Southern US (Garcia et al., 2017). Particularly, the main DTU found in Texas (DTUI) bears a high medical importance as causative agent of Chagasic cardiac disease, highlighting the necessity for further research on this topic in the US (Curtis-Robles et al., 2017b; Garcia et al., 2017).

Previous studies have found a significant change in the microbiome composition of different South American species of Triatominae when challenged with a *Trypanosoma cruzi* infected blood meal (Díaz et al., 2016). Working with wild populations, we collected *Trypanosoma cruzi* positive and negative individuals, which allowed us to assess the correlation between the microbiome and the infection status of *T. protracta* adults. We found that the diversity of the microbiomes in *Trypanosoma cruzi* positive individuals is significantly higher than in the negative ones. This result remarkably agrees with the findings on *Trypanosoma cruzi* challenged *T. brasiliensis, T. sherlocki,* and *P. megistus* (Díaz et al., 2016). However, due to the limited number of available samples, further studies will be necessary to confirm this correlation in natural Triatominae populations.

Arsenophonus triatominarum

A. triatominarum has been described as an intracellular symbiont with strict specificity to various organs and tissues in *Triatoma infestans* (Hypša, 1993). It is therefore interesting to see this bacterium as one of the dominant components of Triatominae gut microbiomes. The presence of *A. triatominarum* in the host guts, and feces, could provide an explanation for the intriguing host distribution of these bacteria. It has been noted that *A. triatominarum* is exclusively bound to Triatominae, but does not show phylogenetic concordance with the host (Sorfová et al., 2008). In fact, the extremely low molecular diversity of the Triatominae bound *Arsenophonus* suggests a recent acquisition and rapid horizontal spread rather than a long time coevolution. While the supposed intracellular location within internal organs was difficult to reconcile with such frequent and host specific horizontal transfers, the occurrence in gut content offers a new possible explanation, such as coprophagy or environmental contamination. Another interesting aspect of *A. triatominarum* being part of the gut microbiome is the possible role of this symbiont. While several members of the genus *Arsenophonus* are known for their profound effect on the host (reproduction distorters or nutritional mutualists; Wilkes et al., 2011; Nováková et al., 2015, 2016), no such phenomenon has so far been observed for *A. triatomiarum*. Considering a possible role of this symbiont, it might be interesting to note that unlike the other two major OTUs, *Enterococcus* and *Acinetobacter*, its relative abundance seems to grow from the first until the third nymphal stages and then drops afterwards. The study of this symbiotic bacterium in the context of the complex and dynamic gut microbiome thus offers a new opportunity to search for its possible effects on the host fitness.

CONCLUSION

The results presented here demonstrate that the composition of Triatominae microbiome is host-specific, maintaining its identity across large geographic areas. Compared to many other insect microbiomes, its overall diversity is low, with 12-17 main bacterial OTUs identified in wild populations of *Triatoma sanguisuga* and *T. protracta*. Surprisingly, one of the dominant taxa is the symbiotic bacterium *Arsenophonus triatominarum*, previously known as an exclusively intracellular symbiont from several Triatominae species. Based on the observations of a laboratory reared colony of *Rhodnius prolixus*, the frequency of the most abundant bacterial genera (*Enterococcus*, *Acinetobacter,* and *Arsenophonus* in this Triatominae species) undergoes significant and consistent changes during the host ontogeny. Finally, a screening of *T. protracta* indicates possible relation between the microbiomes diversity and the occurrence of *Trypanosoma cruzi* in the host gut. This finding highlights the significance of further studies on the microbiomes of NA Triatominae species as a background for developing efficient vector control strategies.

AUTHOR CONTRIBUTIONS

EN and SR-R designed the study. RR, JS, and WR managed the field collections and colony rearing. VS, SR-R and EN generated the data and performed the data analyses. EN, SR-R, and VH drafted the manuscript. All the authors read and contributed to the final text.

ACKNOWLEDGMENTS

We would like to thank Christiane Weirauch who provided us with *T. protracta* samples, Roman Hrdlicka for the care he pays to the Triatominae colonies, and Noah Fierer's laboratory for conducting the sequencing. The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (grant LH-15161 under the program for international collaboration KONTAKT II), and the Grant Agency of the Czech Republic (project no. 18-24707Y).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01167/full#supplementary-material

REFERENCES

Aksoy, E., Telleria, E. L., Echodu, R., Wu, Y., Okedi, L. M., Weiss, B. L., 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. doi: 10.1128/AEM.00079-14

Amino, R., Porto, R. M., Chammas, R., Egami, M. I., and Schenkman, S. (1998). Identification and characterization of a sialidase released by the salivary gland of the hematophagous insect *Triatoma infestans*. *J. Biol. Chem.* 273, 24575–24582. doi: 10.1074/jbc.273.38.24575

Azambuja, P., Feder, D., and Garcia, E. S. (2004). Isolation of *Serratia marcescens* in the midgut of *Rhodnius prolixus*: impact on the establishment of the parasite *Trypanosoma cruzi* in the vector. *Exp. Parasitol.* 107, 89–96. doi: 10.1016/j.exppara.2004.04.007

Azambuja, P., Garcia, E. S., and Ratcliffe, N. A. (2005). Gut microbiota and parasite transmission by insect vectors. *Trends Parasitol.* 21, 568–572. doi: 10.1016/j.pt.2005.09.011

Bern, C., Kjos, S., Yabsley, M. J., and Montgomery, S. P. (2011). *Trypanosoma cruzi* and Chagas' disease in the United States. *Clin. Microbiol. Rev.* 24, 655–681. doi: 10.1128/CMR.00005-11

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26, 266–267. doi: 10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Cirimotich, C. M., Ramirez, J. L., and Dimopoulos, G. (2011). Native microbiota shape insect vector competence for human pathogens. *Cell Host Microbe* 10, 307–310. doi: 10.1016/j.chom.2011.09.006

Crotti, E., Balloi, A., Hamdi, C., Sansonno, L., Marzorati, M., Gonella, E., et al. (2012). Microbial symbionts: a resource for the management of insect-related problems. *Microb. Biotechnol.* 5, 307–317. doi: 10.1111/j.1751-7915.2011.00312.x

Curtis-Robles, R., Hamer, S. A., Lane, S., Levy, M. Z., and Hamer, G. L. (2017a). Bionomics and spatial distribution of triatomine vectors of *Trypanosoma cruzi* in Texas and other southern states, USA. *Am. J. Trop. Med. Hyg.* 98, 113–121. doi: 10.4269/ajtmh.17-0526

Curtis-Robles, R., Snowden, K. F., Dominguez, B., Dinges, L., Rodgers, S., Mays, G., et al. (2017b). Epidemiology and molecular typing of *Trypanosoma cruzi* in naturally-infected hound dogs and associated triatomine vectors in Texas, USA. *PLoS Negl. Trop. Dis.* 11:e0005298. doi: 10.1371/journal.pntd.0005298

Curtis-Robles, R., Wozniak, E. J., Auckland, L. D., Hamer, G. L., and Hamer, S. A. (2015). Combining public health education and disease ecology research: using citizen science to assess Chagas disease entomological risk in Texas. *PLoS Negl. Trop. Dis.* 9:e0004235. doi: 10.1371/journal.pntd.0004235

da Mota, F. F., Marinho, L. P., Moreira, C. J., Lima, M. M., Mello, C. B., Garcia, E. S., et al. (2012). Cultivation-independent methods reveal differences among bacterial gut microbiota in triatomine vectors of Chagas disease. *PLoS Negl. Trop. Dis.* 6:e1631. doi: 10.1371/journal.pntd.0001631

de Fuentes-Vicente, J. A., Gutiérrez-Cabrera, A. E., Flores-Villegas, A. L., Lowenberger, C., Benelli, G., Salazar-Schettino, P. M., et al. (2018). What makes an effective Chagas disease vector? Factors underlying *Trypanosoma cruzi*-triatomine interactions. *Acta Trop.* 183, 23–31. doi: 10.1016/j.actatropica.2018.04.008

Díaz, S., Villavicencio, B., Correia, N., Costa, J., and Haag, K. L. (2016). Triatomine bugs, their microbiota and *Trypanosoma cruzi*: asymmetric responses of bacteria to an infected blood meal. *Parasit. Vectors* 9:636. doi: 10.1186/s13071-016-1926-2

Duguma, D., Hall, M. W., Rugman-Jones, P., Stouthamer, R., Terenius, O., Neufeld, J. D., et al. (2015). Developmental succession of the microbiome of *Culex* mosquitoes. *BMC Microbiol.* 15:140. doi: 10.1186/s12866-015-0475-8

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Edwards, M. S., Stimpert, K. K., and Montgomery, S. P. (2017). Addressing the challenges of Chagas disease: an emerging health concern in the United States. *Infect. Dis. Clin. Pract.* 25:118. doi: 10.1097/IPC.0000000000000512

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*, *2nd Edn*. Thousand Oaks, CA: Sage Publications. Available online at: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Galvão, C., Carcavallo, R., Rocha, D. D. S., and Jurberg, J. (2003). A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa* 202, 1–36. doi: 10.11646/zootaxa.202.1.1

Garcia, E. S., Castro, D. P., Figueiredo, M. B., and Azambuja, P. (2010). Immune homeostasis to microorganisms in the guts of triatomines (Reduviidae): a review. *Mem. Inst. Oswaldo Cruz* 105, 605–610. doi: 10.1590/S0074-02762010000500001

Garcia, M. N., Burroughs, H., Gorchakov, R., Gunter, S. M., Dumonteil, E., Murray, K. O., et al. (2017). Molecular identification and genotyping of *Trypanosoma cruzi* DNA in autochthonous Chagas disease patients from Texas, USA. *Infect. Genet. Evol.* 49, 151–156. doi: 10.1016/j.meegid.2017.01.016

Gimonneau, G., Tchioffo, M. T., Abate, L., Boissière, A., Awono-Ambéné, P. H., Nsango, S. E., et al. (2014). Composition of *Anopheles coluzzii* and *Anopheles gambiae* microbiota from larval to adult stages. *Infect. Genet. Evol.* 28, 715–724. doi: 10.1016/j.meegid.2014.09.029

Gourbière, S., Dorn, P., Tripet, F., and Dumonteil, E. (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity* 108, 190–202. doi: 10.1038/hdy.2011.71

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gumiel, M., da Mota, F. F., Rizzo Vde, S., Sarquis, O., de Castro, D. P., Lima, M. M., et al. (2015). Characterization of the microbiota in the guts of *Triatoma brasiliensis* and *Triatoma pseudomaculata* infected by *Trypanosoma cruzi* in natural conditions using culture independent methods. *Parasit. Vectors* 8:245. doi: 10.1186/s13071-015-0836-z

Gunter, S. M., Murray, K. O., Gorchakov, R., Beddard, R., Rossmann, S. N., Montgomery, S. P., et al. (2017). Likely autochthonous transmission of *Trypanosoma cruzi* to humans, south central Texas, USA. *Emerg. Infect. Dis.* 23, 500–503. doi: 10.3201/eid2303.161157

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W., and Nadler, S. A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265, 1087–1090. doi: 10.1126/science.8066445

Hegde, S., Rasgon, J. L., and Hughes, G. L. (2015). The microbiome modulates arbovirus transmission in mosquitoes. *Curr. Opin. Virol.* 15, 97–102. doi: 10.1016/j.coviro.2015.08.011

Hernández, C., Salazar, C., Brochero, H., Teherán, A., Buitrago, L. S., Vera, M., et al. (2016). Untangling the transmission dynamics of primary and secondary vectors of *Trypanosoma cruzi* in Colombia: parasite infection, feeding sources and discrete typing units. *Parasit. Vectors* 9:620. doi: 10.1186/s13071-016-1907-5

Husseneder, C., Park, J. S., Howells, A., Tikhe, C. V., and Davis, J. A. (2017). Bacteria associated with *Piezodorus guildinii* (Hemiptera: Pentatomidae), with special reference to those transmitted by feeding. *Environ. Entomol.* 46, 159–166. doi: 10.1093/ee/nvw112

Hypša, V. (1993). Endosytobionts of *Triatoma infestans*: distribution and transmission. *J. Invertebr. Pathol.* 61, 32–38. doi: 10.1006/jipa.1993.1006

Hypša, V., and Aksoy, S. (2003). Phylogenetic characterization of two transovarially transmitted endosymbionts of the bedbug *Cimex lectularius* (Heteroptera: Cimicidae). *Insect Mol. Biol.* 6, 301–304. doi: 10.1046/j.1365-2583.1997.00178.x

Jupatanakul, N., Sim, S., and Dimopoulos, G. (2014). The insect microbiome modulates vector competence for arboviruses. *Viruses* 6, 4294–4313. doi: 10.3390/v6114294

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. doi: 10.1093/bioinformatics/bts199

Klepzig, K. D., Adams, A. S., Handelsman, J., and Raffa, K. F. (2009). Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environ. Entomol.* 38, 67–77. doi: 10.1603/022.038.0109

Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47, 583–621. doi: 10.1080/01621459.1952.10483441

Kwong, W. K., Engel, P., Koch, H., and Moran, N. A. (2014). Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11509–11514. doi: 10.1073/pnas.1405838111

Leonhardt, S. D., and Kaltenpoth, M. (2014). Microbial communities of three sympatric Australian stingless bee species. *PLoS ONE* 9:e10571 doi: 10.1371/journal.pone.0105718

McPhatter, L., Roachell, W., Mahmood, F., Hoffman, L., Lockwood, N., Osuna, A., et al. (2012). Vector surveillance to determine species composition and occurrence of *Trypanosoma cruzi* at three military installations in San Antonio, Texas. *US Army Med. Dep. J.* 3–12, 12–21.

Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., et al. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14936–14941. doi: 10.1073/pnas.1506226112

Minard, G., Mavingui, P., and Moro, C. V. (2013). Diversity and function of bacterial microbiota in the mosquito holobiont. *Parasit. Vectors* 6:146. doi: 10.1186/1756-3305-6-146

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69, 89–107. doi: 10.1007/BF00038690

Montgomery, S. P., Starr, M. C., Cantey, P. T., Edwards, M. S., and Meymandi, S. K. (2014). Neglected parasitic infections in the United States: Chagas disease. *Am. J. Trop. Med. Hyg.* 90, 814–818. doi: 10.4269/ajtmh.13-0726

Moser, D. R., Kirchhoff, L. V., and Donelson, J. E. (1989). Detection of *Trypanosoma cruzi* by DNA amplification using the polymerase chain reaction. *J. Clin. Microbiol.* 27, 1477–1482.

Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., and Fukatsu, T. (2014). Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proc. Nat. Acad. Sci. U.S.A.* 111, 10257–10262. doi: 10.1073/pnas.1409284111

Nováková, E., Husník, F., Šochová, E., and Hypša, V. (2015). *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Appl. Environ. Microbiol.* 81, 6189–6199. doi: 10.1128/AEM.01487-15

Nováková, E., Hypša, V., Nguyen, P., Husník, F., and Darby, A. C. (2016). Genome sequence of Candidatus *Arsenophonus lipopteni*, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *Stand. Genomic Sci.* 11:72. doi: 10.1186/s40793-016-0195-1

Novakova, E., Woodhams, D. C., Rodríguez-Ruano, S. M., Brucker, R. M., Leff, J. W., Maharaj, A., et al. (2017). Mosquito microbiome dynamics, a background for prevalence and seasonality of West Nile Virus. *Front. Microbiol.* 8:526. doi: 10.3389/fmicb.2017.00526

Nunes-da-Fonseca, R., Berni, M., Tobias-Santos, V., Pane, A., and Araujo, H. M. (2017). *Rhodnius prolixus*: from classical physiology to modern developmental biology. *Genesis* 55:e22995. doi: 10.1002/dvg.22995

Núñez, J. A., and Lazzari, C. R. (1990). Rearing of *Triatoma infestans* Klug (Het., Reduviidae) in the absence of a live host. *J. Appl. Entomol.* 109, 87–92. doi: 10.1111/j.1439-0418.1990.tb00023.x

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013). *Vegan: Community Ecology Package*. The Comprehensive R Archive Network (CRAN) Available online at: http://CRAN.R-project.org/package=vegan.

Oliver, K. M., and Martinez, A. J. (2014). How resident microbes modulate ecologically-important traits of insects. *Curr. Opin. Insect Sci.* 4, 1–7. doi: 10.1016/j.cois.2014.08.001

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–D596. doi: 10.1093/nar/gks1219

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna. Available online at: http://www.R-project.org/

Saldaña, M. A., Hegde, S., and Hughes, G. L. (2017). Microbial control of arthropod-borne disease. *Mem. Inst. Oswaldo Cruz* 112, 81–93. doi: 10.1590/0074-02760160373

Sorfová, P., Skeríková, A., and Hypsa, V. (2008). An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of *Arsenophonus triatominarum*. *Syst. Appl. Microbiol.* 31, 88–100. doi: 10.1016/j.syapm.2008.02.004

Sudakaran, S., Salem, H., Kost, C., and Kaltenpoth, M. (2012). Geographical and ecological stability of the symbiotic mid-gut microbiota in European firebugs, *Pyrrhocoris apterus* (Hemiptera, Pyrrhocoridae). *Mol. Ecol.* 21, 6134–6151. doi: 10.1111/mec.12027

Taracena, M. L., Oliveira, P. L., Almendares, O., Umaña, C., Lowenberger, C., Dotson, E. M., et al. (2015). Genetically modifying the insect gut microbiota to control Chagas disease vectors through systemic RNAi. *PLoS Negl. Trop. Dis.* 9:e0003358. doi: 10.1371/journal.pntd.0003358

Weiss, B., and Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends Parasitol.* 27, 514–522. doi: 10.1016/j.pt.2011.05.001

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5:27. doi: 10.1186/s40168-017-0237-y

WHO (2015). Chagas disease in Latin America: an epidemiological update based on 2010 estimates. *Releve Epidemiol. Hebd.* 90, 33–43.

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag Available online at: http://ggplot2.org.

Wilkes, T. E., Duron, O., Darby, A. C., Hypša, V., Nováková, E., and Hurst, G. D. D. (2011). "The genus Arsenophonus," in *Manipulative Tenants: Bacteria Associated With Arthropods,* eds E. Zchori-Fein and K. Bourtzis, (Boca Raton, FL: CRC Press), 520–528.

Wozniak, E. J., Lawrence, G., Gorchakov, R., Alamgir, H., Dotson, E., Sissel, B., et al. (2015). The biology of the Triatomine bugs native to south central Texas and assessment of the risk they pose for autochthonous Chagas disease exposure. *J. Parasitol.* 101, 520–528. doi: 10.1645/15-748

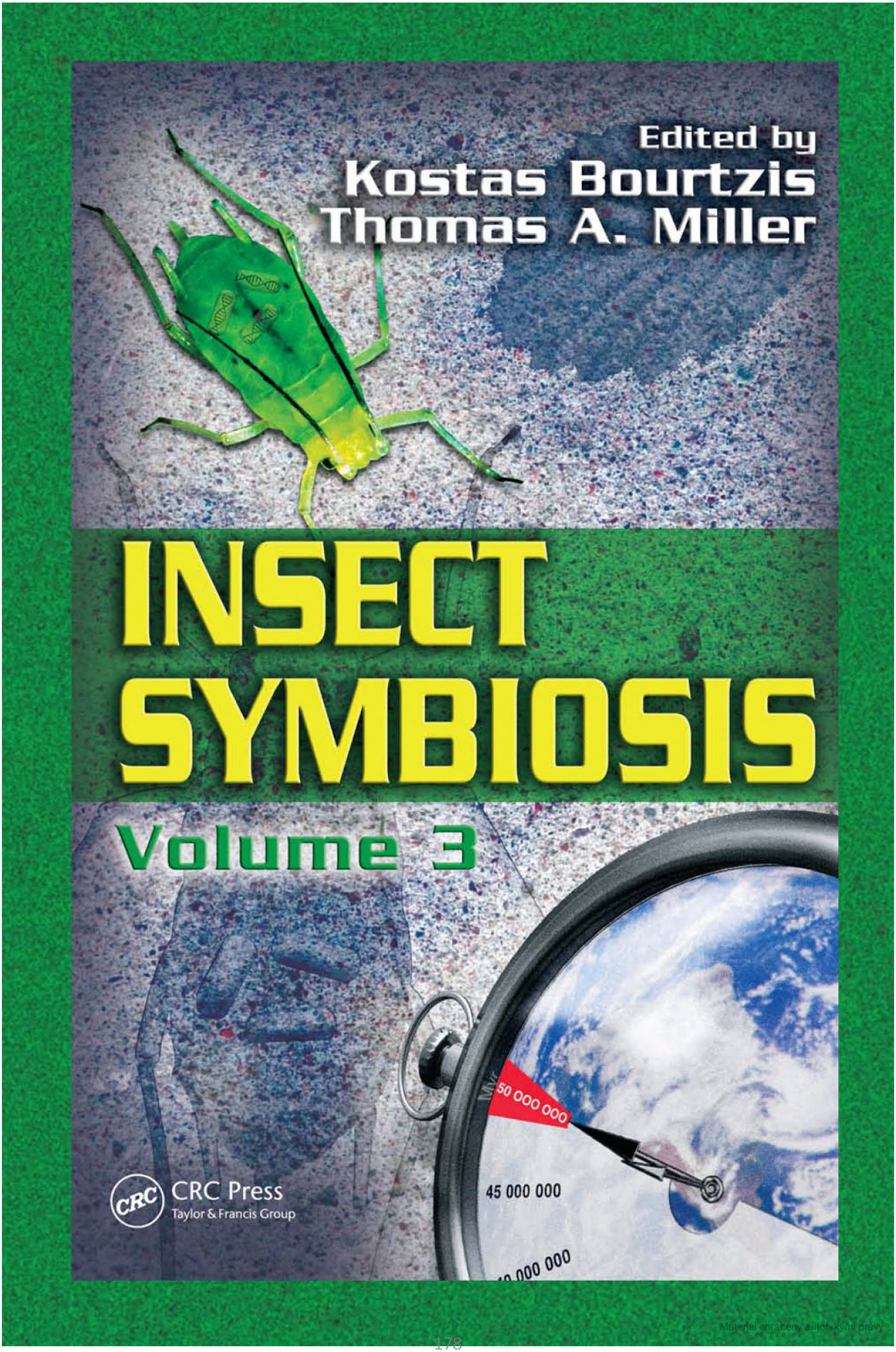
Zhang, G., Browne, P., Zhen, G., Johnston, A., Cadillo-Quiroz, H., and Franz, N. (2017). Endosymbiont diversity and evolution across the weevil tree of life. *bioRxiv*, 171181. doi: 10.1101/171181

Zolnik, C. P., Prill, R. J., Falco, R. C., Daniels, T. J., and Kolokotronis, S. O. (2016). Microbiome changes through ontogeny of a tick pathogen vector. *Mol. Ecol.* 25, 4963–4977. doi: 10.1111/mec.13832

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Book Chapter 1:

Hypša, V., & **Nováková, E.** (2009). Insect symbionts and molecular phylogenetics. *Insect symbiosis*, *3*, 1-31.



CONTEMPORARY TOPICS in ENTOMOLOGY SERIES

THOMAS A. MILLER Editor

INSECT SYMBIOSIS

Volume 3

Edited by

Kostas Bourtzis

Thomas A. Miller

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Cover: The genome sequences of the *Buchnera aphidicola* symbionts of *Schizaphis graminum* and *Acyrthosiphon pisum* have been determined. A comparison of the two genomes shows that no rearrangements or gene acquisition have occurred in the past 50 to 70 million years, despite the high levels of nucleotide-sequence divergence. This is the first time that whole-genome evolution for microbes has been calibrated with respect to time. The analysis has shown that *B. aphidicola* have the most stable genomes characterized to date. (Photograph courtesy of Ola Lundström, Department of Molecular Evolution, Uppsala University, Uppsala, Sweden.)

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-6410-0 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Insect symbiosis vol. 3 / edited by Kostas Bourtzis and Thomas A. Miller.
p. cm. -- (Contemporary topics in entomology)
Includes bibliographical references (p.).
ISBN 978-1-4200-6410-0
1. Insects--Ecology. 2. Symbiosis. I. Bourtzis, Kostas. II.Miller, Thomas A. III. Series.

QL496.4 .I57 2003
595.71785--dc21 2002038796

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface xi

About the Editors xiii

Contributors xv

Chapter one Insect symbionts and molecular phylogenetics 1
Václav Hypša and Eva Nováková

Chapter two Self–nonself recognition in symbiotic interactions 33
Otto Schmidt

Chapter three Is symbiosis evolution influenced by the pleiotropic role of programmed cell death in immunity and development? 57
Fabrice Vavre, Natacha Kremer, Bart A. Pannebakker, Benjamin Loppin, and Patrick Mavingui

Chapter four Pleiotropy of adaptative genes: how insecticide resistance genes mediate outcomes of symbiosis 77
Olivier Duron and Mylène Weill

Chapter five Capsule-transmitted obligate gut bacterium of plataspid stinkbugs: a novel model system for insect symbiosis studies 95
Takema Fukatsu and Takahiro Hosokawa

Chapter six Endosymbiont that broadens food plant range of host insect 123
Tsutomu Tsuchida, Ryuichi Koga, and Takema Fukatsu

Chapter seven Insect–bacterium mutualism without vertical transmission 143
Yoshitomo Kikuchi and Takema Fukatsu

Chapter eight Mutualism revealed by symbiont genomics and bacteriocyte transcriptomics 163
Atsushi Nakabachi

Chapter nine Endosymbionts of lice 205
M. Alejandra Perotti, Ewen F. Kirkness, David L. Reed, and Henk R. Braig

Chapter ten Symbiotic *Rickettsia* 221
Henk R. Braig, Bryan D. Turner, and M. Alejandra Perotti

Chapter eleven Structure and function of the bacterial community associated with the Mediterranean fruit fly 251
Adi Behar, Michael Ben-Yosef, Carol R. Lauzon, Boaz Yuval, and Edouard Jurkevich

Chapter twelve Feminizing *Wolbachia* and the evolution of sex determination in isopods 273
Didier Bouchon, Richard Cordaux, and Pierre Grève

Chapter thirteen *Wolbachia*-induced sex reversal in Lepidoptera 295
Satoko Narita and Daisuke Kageyama

Chapter fourteen *Wolbachia* and *Anopheles* mosquitoes 321
Jason L. Rasgon

Chapter fifteen Bacterial symbionts in *Anopheles* spp. and other mosquito vectors 329
Guido Favia, Daniele Daffonchio, Luciano Sacchi, Alberto Alma, and Claudio Bandi

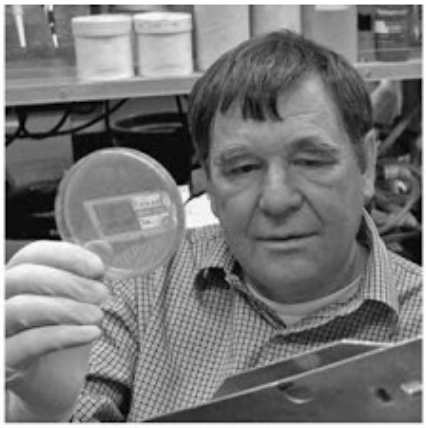
Chapter sixteen Symbiotic microorganisms in leafhopper and planthopper vectors of phytoplasmas in grapevine 339
Alberto Alma, Luciano Sacchi, Guido Favia, Claudio Bandi, and Daniele Daffonchio

Chapter seventeen Paratransgenesis in termites 361
Claudia Husseneder and Rachael E. Collier

Chapter eighteen Insect facultative symbionts: biology, culture, and genetic modification 377
Mauricio Pontes, Kari Smith, Wendy Smith, and Colin Dale

Index 397

Dedication to Paul Baumann



I still remember very clearly my first conversation with Paul Baumann. He had phoned to ask if I might be interested in collaborating on a study on the bacterial endosymbionts of aphids; the year was 1990. His excitement was infectious. I had read much of Paul Buchner's book as a graduate student but had never considered working on symbionts and knew almost nothing of microbiology.

During the earlier parts of his career, Paul had already contributed major work on topics related to marine bacteria and *Bacillus* pathogens of mosquitoes. Now he saw the possibility of understanding diversity of noncultivable symbionts in a firm molecular phylogenetic context and, further, to use molecular methods to explore symbiont functions. When I met him, Paul already had produced the first sequence-based characterization of obligate endosymbionts, work that resulted in the unexpected finding that the primary endosymbionts, which he named *Buchnera aphidicola*, were rather closely related to *Escherichia coli* (Unterman and Baumann, 1989). He realized that a close relationship to *E. coli* was a godsend for further investigations on functional capabilities of the symbiont. At that time, few genes were sequenced for any organisms, and homology to the best-studied model organism made it more likely that symbiont genes could be identified, cloned, and sequenced as a step toward understanding functional capabilities of these noncultivable organisms. Displaying a habit that I was to discover was usual, Paul had done exhaustive reading of the relevant literature, including papers on nutrition of aphids that suggested a role of symbionts in nutrition. By bringing new molecular data to the understanding of symbionts, he opened a new frontier in symbiosis studies and in the broader understanding of how microbes interface with multicellular eukaryotes.

Even when his results were groundbreaking, and even though his own enthusiasm was always immense, Paul had been consistently understated in how he presented results in publications and in speaking. He maintained an old-fashioned faith in his fellow scientists: that they would recognize an important result on its own merits and that "hype" would only distract (or insult) worthy readers. The first paper on codiversification of *Buchnera* and aphids (Munson et al., 1991) was a very important one, the first of many studies from many investigators to show that the phylogeny of bacteriome-associated symbionts mirrors that of their hosts. These matching phylogenies of host and symbiont are the strongest evidence for the antiquity of the associations, implying that the symbionts descend from an ancient infection predating the origin of the host group (~150 My in the case of aphids).

But this result was really a single sentence in the discussion of a paper in *Journal of Bacteriology*, without a figure. Often, this important result is attributed to a later paper (Moran et al., 1993), which expanded it and added a calibrated rate of evolution for the bacteria.

Paul, working with his wife, Linda, and others in his group at Davis, continued to methodically clone and sequence fragments identified on the basis of homology to *E. coli*, focusing on the genes underlying the biosynthesis of essential amino acids. The size of the *Buchnera* genome had earlier been claimed to be fivefold greater than that of *E. coli*, so, when the age of genomics dawned in 1995 with the full genome sequence of *Haemophilus influenzae*, *Buchnera* did not seem an obvious target for genomic sequencing (this was later shown to be wrong, with the *Buchnera* genome only one-seventh that of *E. coli*). By the time the first genome of *Buchnera* was published in 2000, Paul had characterized many *Buchnera* genes, including those encoding numerous enzymes in amino acid biosynthesis, and he had discovered that the genes underlying the rate-limiting step of tryptophan synthesis were amplified and located on an unusual plasmid, an apparent adaptation of the symbiont to better supply its host with this required nutrient (Lai and Baumann, 1994).

While understanding symbiont function has been one of Paul's primary aims, exploring unknown diversity has been the other. Between 1990 and 2005, he catalogued, sometimes collaborating with me, the remarkable assortment of symbiont types within most of the major insect groups related to aphids, including whiteflies, psyllids, and mealybugs (summarized in Baumann, 2005).

Part of why he was able to put the study of symbiosis on a firm foundation using molecular biology was that he avoided vague speculations and adhered to meticulous standards for data collection and quality. He believed in getting the facts right. As an illustration, when he once discovered minor errors in DNA sequences previously submitted to GenBank (at a time when bases were still being called manually from autorads of sequencing gels and when unresolved and erroneous positions were frequent in sequence databases), he resequenced the same templates and corrected the submissions, even though the papers were already published and even though there were no changes to the conclusions. His aversion to sloppiness was a force helping to set standards for the emerging field of molecular biology of symbiosis.

Although a perfectionist, his science was fueled by a sense of adventure and an attraction to exploring the unknown. In fact, much was unknown, including much that, thanks in large part to his efforts, is now known and taken for granted by students and others just starting in the field of symbiosis studies. It is now widely accepted that symbionts extend deep into the evolutionary history of insects and other eukaryotes and that we can explore their functions using knowledge of genetics and molecular biology derived from model systems. Paul and Linda's work is a major reason for these insights and others.

Paul and Linda retired in 2005. He is pursuing long-term interests in music, photography, and history (all areas in which he has remarkable expertise and characteristic enthusiasm). Symbiosis research, now hugely expanded compared to when he started, owes much to his pioneering vision and high standards.

Nancy A. Moran
University of Arizona

Some Citations from Paul Baumann's Work

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

Lai, C.Y., Baumann, L., and Baumann, P. (1994). Amplification of trpEG—adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 3819–3823.

Moran, N.A., Munson, M.A., Baumann, P., and Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lon. B–Biolog. Sci.* **253:** 167–171.

Munson, M.A., Baumann, P., Clark, M.A., Baumann, L., Moran, N.A., Voegtlin, D.J., and Campbell, B.C. (1991). Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of 4 aphid families. *J. Bacteriol.* **173:** 6321–6324.

Unterman, B.M., Baumann, P., and McLean, D.L. (1989). Pea aphid symbiont relationships established by analysis of 16S ribosomal RNAs. *J. Bacteriol.* **171**: 2970–2974.

Preface to Insect Symbiosis, Volume 3

Volume 1 of this series contained a foreword by the noted science writer Carl Zimmer and was dedicated to Professor Frank Richards for his pioneering work in paratransgenesis and symbiotic control. The second volume had a foreword from Lynn Margulis, the noted symbiosis pioneer, and an obituary of Professor Hajime Ishikawa, a pioneer in insect symbiosis, written by his outstanding student Takema Fukatsu. This third volume has this preface and a dedication to the pioneering work of Paul Baumann in insect symbiosis, written by his long-time collaborator, Nancy Moran.

The field of insect symbiosis continues to grow. As we write this, Kostas leads a consortium of European colleagues at the inaugural meeting of the symbiosis group to initiate funding in the European Union for research on symbiosis to discover new tools for pest and disease control in agriculture and medicine (Brussels, Belgium, March 2008), and Thomas helps organize a meeting called by the Ministry of Agriculture, Rabat, Morocco, aimed at finding new tools of biotechnology for control of desert locust, *Schistocerca gregaria*. Symbiosis offers one bright hope for a breakthrough in this historically difficult pest problem.

The present volume contains reports from outstanding laboratories across the field of insect symbiosis and includes work suggesting or hinting at practical applications in mosquitoes, tephritids, and termites. There are more examples in this volume of the influence of symbionts on the biology of insects.

About the Editors

Kostas Bourtzis, Ph.D., is Associate Professor of Molecular Biology and Biochemistry in the Department of Environmental and Natural Resources Management, University of Ioannina, Greece. His research interests include *Wolbachia*-mediated cytoplasmic incompatibility in *Drosophila*, agricultural insect pests and disease vectors; genetic manipulation of *Wolbachia*; molecular mechanism of cytoplasmic incompatibility; *Wolbachia* genomics; and the use of endosymbiotic bacteria including *Wolbachia* as a tool for the development of new, environmentally friendly approaches for the control of arthropods of medical and agricultural importance. His group has recently shown that *Wolbachia*-induced cytoplasmic incompatibility can be used as a means to suppress insect pest populations.

Thomas A. Miller, Ph.D., is Professor of Entomology at the University of California, Riverside, where he teaches insect physiology and insect toxicology. A B.A. in physics influenced his earlier work in insect neurophysiology, including the development of electronic transducers for measuring small muscle forces. This was followed by contributions to mode of action of insecticides and developing methods of measuring resistance in cotton pests, which again involved the development of unique electromechanical devices to record insect activity in the field. Dr. Miller led the successful effort to genetically transform the pink bollworm and provided the opportunity to employ a conditional lethal strain for use in area-wide eradication. All of these achievements were done with students, colleagues, and collaborators. Most recently, Dr. Miller has been seeking ways of applying principles of biotechnology to pest and disease control, again with a host of collaborators.

Contributors

Alberto Alma
Dipartimento di Valorizzazione e
Protezione delle Risorse Agroforestali
Università degli Studi di Torino
Torino, Italy

Claudio Bandi
Dipartimento di Patologia Animale
Igiene e Sanità Pubblica Veterinaria
Università degli Studi di Milano
Milano, Italy

Adi Behar
Departments of Microbiology and
Entomology
Hebrew University of Jerusalem
Rehovot, Israel

Michael Ben-Yosef
Department of Microbiology
Hebrew University of Jerusalem
Rehovot, Israel

Didier Bouchon
Laboratoire Ecologie Evolution Symbiose
Université de Poitiers
Poitiers, France

Henk R. Braig
School of Biological Sciences
Bangor University
Bangor, UK

Rachael E. Collier
Department of Entomology
Louisiana State University Agricultural
Center
Baton Rouge, LA

Richard Cordaux
Laboratoire Ecologie Evolution Symbiose
Université de Poitiers
Poitiers, France

Daniele Daffonchio
Dipartimento di Scienze e Tecnologie
Alimentari e Microbiologiche
Università degli Studi di Milano
Milano, Italy

Colin Dale
Department of Biology
University of Utah
Salt Lake City, UT

Olivier Duron
Institut des Sciences de l'Evolution
Equipe Génétique de l'Adaptation
Université Montpellier
Montpellier, France
and
Department of Biology
University College London
London, UK

Guido Favia
Dipartimento di Medicina Sperimentale e
Sanità Pubblica
Università degli Studi di Camerino
Camerino, Italy

Takema Fukatsu
National Institute of Advanced Industrial
Science and Technology
Tsukuba, Ibaraki, Japan

Pierre Grève
Laboratoire Ecologie Evolution Symbiose
Université de Poitiers
Poitiers, France

Takahiro Hosokawa
National Institute of Advanced Industrial Science and Technology
Tsukuba, Ibaraki, Japan

Claudia Husseneder
Department of Entomology
Louisiana State University Agricultural Center
Baton Rouge, LA

Václav Hypša
Faculty of Science
University of South Bohemia and Institute of Parasitology
Branišovská, České Budějovice, Czech Republic

Edouard Jurkevich
Department of Microbiology
Hebrew University of Jerusalem
Rehovot, Israel

Daisuke Kageyama
National Institute of Agrobiological Sciences
Tsukuba, Ibaraki, Japan

Yoshitomo Kikuchi
National Institute of Advanced Industrial Science and Technology
Tsukuba, Ibaraki, Japan

Ewen F. Kirkness
The Institute for Genomic Research
Rockville, MD

Ryuichi Koga
National Institute of Advanced Industrial Science and Technology
Tsukuba, Ibaraki, Japan

Natacha Kremer
Laboratoire de Biométrie et Biologie Evolutive
Université de Lyon
Lyon, France

Carol R. Lauzon
Department of Biological Sciences
California State University
Hayward, CA

Benjamin Loppin
Centre de Génétique Moléculaire et Cellulaire
Université de Lyon
Lyon, France

Patrick Mavingui
Université de Lyon
Lyon, France

Atsushi Nakabachi
Frontier Research System
Wako, Saitama, Japan

Satoko Narita
National Institute of Agrobiological Sciences
Tsukuba, Ibaraki, Japan

Eva Nováková
Faculty of Science
University of South Bohemia and Institute of Parasitology
Branišovská, České Budějovice, Czech Republic

Bart A. Pannebakker
Institute of Evolutionary Biology
School of Biological Sciences
University of Edinburgh
Edinburgh, Scotland, UK

M. Alejandra Perotti
School of Biological Sciences
University of Reading
Reading, UK
Mauricio Pontes
Department of Biology
University of Utah
Salt Lake City, UT

Jason L. Rasgon
W. Harry Feinstone Department of Molecular Microbiology and Immunology, and the Johns Hopkins Malaria Research Institute
Bloomberg School of Public Health
Johns Hopkins University
Baltimore, MD

David L. Reed
Florida Museum of Natural History
University of Florida
Gainesville, FL

Luciano Sacchi
Dipartimento di Biologia Animale
Università degli Studi di Pavia
Pavia, Italy

Otto Schmidt
Insect Molecular Biology
University of Adelaide
Glen Osmond, Australia

Kari Smith
Department of Biology
University of Utah
Salt Lake City, UT

Wendy Smith
Department of Biology
University of Utah
Salt Lake City, UT

Tsutomu Tsuchida
National Institute of Advanced Industrial Science and Technology
Tsukuba, Ibaraki, Japan

Bryan D. Turner
Department of Forensic Science and Drug Monitoring
King's College London
London, UK

Fabrice Vavre
Laboratoire de Biométrie et Biologie Evolutive
Université de Lyon
Lyon, France

Mylène Weill
Institut des Sciences de l'Evolution
Equipe Génétique de l'Adaptation
Université Montpellier
Montpellier, France

Boaz Yuval
Department of Entomology
Hebrew University of Jerusalem
Rehovot, Israel

chapter one

Insect symbionts and molecular phylogenetics

Václav Hypša and Eva Nováková

Contents

The significance of molecular phylogenetics for insect symbiont research 1
P-symbionts and S-symbionts: old terms in a new phylogenetic framework 5
Composition of symbiotic fauna: transitions, losses, and replacements 7
P-symbionts: a touchstone of molecular phylogenetics 8
Nonhomogeneous models 8
Multigene approach 10
Genome structure 11
Evolutionary implications of P-symbionts monophyly/polyphyly 12
Coevolutionary reconstructions in symbiotic associations 13
Tree congruence 13
Compatibility of host and symbiont data 14
Tree confidence 15
Molecular clock and calibration of evolutionary rates 16
Specificity without coevolution 16
Low-level phylogeny 18
Intragenomic heterogeneity 19
Diversity and sample 20
Conclusion 22
Acknowledgments 22
References 23

The significance of molecular phylogenetics for insect symbiont research

Investigation into prokaryote evolution is one of the areas of biology most deeply affected by the advent of molecular phylogenetics (Pace, 1997). In the field of insect–bacteria symbiosis, this new approach not only generated an immense amount of knowledge, but also fundamentally changed the whole framework of this field: bacterial symbionts are now mostly viewed and analyzed in terms of *phylogenetic lineages*. This situation contrasts with the premolecular era, when morphological and functional characterization provided the only base for symbiont classification. Until the late 1980s, studies on insect bacterial symbionts were either purely descriptive or focused on functional rather than evolutionary aspects (Buchner, 1965; Hill and Campbell, 1973; Griffiths and Beck,

1974; Nogge, 1978; Ishikawa, 1982; Eberle and McLean, 1983); when tackling evolutionary issues at all, they were largely speculative and could rely on only indirect evidence (Buchner, 1965). The new phylogeny-based way of looking at the symbionts allowed for addressing many previously inapproachable issues, such as the origin of various symbiotic lineages in frame of the whole prokaryote diversity, degree of vertical versus horizontal transfers, or continuous acquisitions and losses of various symbiotic lineages by their hosts. On the other hand, this new approach introduced into the symbiosis field a variety of methodological problems related to the intricacy of phylogenetic inference and coevolutionary reconstructions.

Within less than two decades, a substantial number of analyses have been performed and published. From a formal point of view, this process resulted in identification and taxonomic description of many symbiotic lineages (Table 1.1). More importantly, several interesting patterns began to emerge from the accumulated data. In Figure 1.1, we provide a rough picture of insect–symbiont diversity mapped onto a phylogenetic background. Because there is little agreement on the overall phylogenetic arrangement of bacterial symbionts, this topology is meant as a reference scheme for the following discussions rather than as an attempt to achieve consensus of the current views. Despite the high degree of uncertainty and subjectivity, it demonstrates several important facts. At broad phylogenetic scale, the expected polyphyly of bacterial symbionts was well established. It is now clear that symbiotic lineages have originated multiple times within many groups of prokaryotes. On the other hand, hot spots giving rise to multiple symbiotic lineages can be identified within some taxa. For example, many of the major symbiotic lineages seem to stem from a particularly prominent spot within γ-Proteobacteria (the node P in Figure 1.1). This issue, however, remains highly controversial and will be discussed later.

Another outcome of phylogenetic studies is the apparent versatility of bacterial lineages. Closely related bacteria can rapidly evolve into a highly specialized mutualist as well as parasites/commensals loosely associated with their hosts (nodes P, A, and S in Figure 1.1 show such diverse groups). This versatility has even been experimentally demonstrated by functional replacement of obligate long-established mutualists with other symbiotic bacteria (Koga et al., 2003). Several recent studies indicate that at least some of the symbiotic lineages previously believed to have originated by only a single and unique transition from a free-living bacterium to symbiosis, may in fact represent abundant and widely distributed symbionts (e.g., the genera *Sodalis* and *Arsenophonus*). An increasing tempo of identification and characterization of new symbiotic lineages will certainly lead to an even more complex picture and therefore higher demands on coevolutionary reconstruction and its evolutionary interpretation. Consequently, the methods originally restricted to parasitological research, such as tree-based and data-based coevolutionary reconstructions, have begun to penetrate the field of symbiosis.

In this chapter, we highlight the main questions and methodological problems that arose from the current view on insect–bacteria symbiosis. In the first section, we show that the phylogenetic information accumulated during the last years of molecular research has changed some of the previous conceptions of the field and puts insect–bacteria symbiosis into an entirely new perspective. In this context, we highlight the importance of unprejudiced and rigorous interpretation of phylogenetic relationships for evolutionary inferences. In the following sections, we focus on two major methodological problems, namely, the inference of phylogenetic signal from degenerated symbiotic genomes and the reconstruction of coevolutionary history between insects and their symbionts.

Table 1.1 List of Symbiotic Bacteria Included in the Tree in Figure 1.1

Taxon Designation	Classification	Host	References
Arsenophonus sp.	Gamma	Various arthropods	(Gherna et al., 1991; Hypša and Dale, 1997; Thao and Baumann, 2004b)
Baumannia cicadellinicola	Gamma	Various spp. of sharpshooters	(Moran et al., 2003)
*Bemisia tabaci**	Gamma	**	(Zchori-Fein and Brown, 2002)
***Blochmania* sp.**	Gamma	Various spp. of carpenter ants	(Schröder et al., 1996; Sauer et al., 2000)
Buchnera aphidicola	Gamma	Aphids	(Munson et al., 1991b)
Carsonella ruddii	Gamma	Psyllids	(Spaulding and von Dohlen, 1998; Clark et al., 2001; Nakabachi et al., 2006)
*Cimex lectularius**	Gamma	**	(Hypša and Aksoy, 1997)
*Columbicola columbae**	Gamma	**	(Fukatsu et al., 2007)
*Craterina malbae**	Gamma	**	(Nováková and Hypša, 2007)
*Euscilidius variegatus**	Gamma	**	(Campbell and Purcell, 1993)
Hamiltonella defensa	Gamma	Various spp. of aphids and psyllids	(Moran et al., 2005a)
Hematomizus elephantis*	Gamma	**	(Hypša and Křížek, 2007)
Hematopinus* sp.	Gamma	**	(Hypša and Křížek, 2007)
Ishikawaella capsulata	Gamma	Various spp. of stinkbugs	(Hosokawa et al., 2006)
Mealybugs A*	Gamma	Various spp. of mealybugs	(Thao et al., 2002)
Mealybugs C	Gamma	Various spp. of mealybugs	(Thao et al., 2002)
Mealybugs D*	Gamma	Various spp. of mealybugs	(Thao et al., 2002)
Mealybugs E*	Gamma	Various spp. of mealybugs	(Thao et al., 2002)
Nardonella	Gamma	Various spp. of weevils	(Lefevre et al., 2004)
Polyplax sp.*	Gamma	**	(Hypša and Křížek, 2007)
Portiera aleyrodidarum	Gamma	Whiteflies	(Thao and Baumann, 2004a)
*Pseudolynchia canariensis**	Gamma	**	(Dale et al., 2006)

Continued.

Table 1.1 List of Symbiotic Bacteria Included in the Tree in Figure 1.1 (*Continued*)

Taxon Designation	Classification	Host	References
Psyllids 1*	Gamma	Various spp. of psyllids	(Thao et al., 2000a)
Psyllids 2*	Gamma	Various spp. of psyllids	(Thao et al., 2000a)
Regiella insecticola	Gamma	Various spp. of aphids	(Moran et al., 2005a)
Riesia pediculicola	Gamma	Primate lice	(Allen et al., 2007)
Serratia symbiotica	Gamma	Various spp. of aphids	(Moran et al., 2005a)
Sitophilus* sp.	Gamma	**	(Lefevre et al., 2004)
Sodalis glossinidius	Gamma	*Glossina* sp.	(Dale and Maudlin, 1999; Dale et al., 2001; Toh et al., 2006)
Trichobius* sp.	Gamma	Various spp. of bat flies	(Trowbridge et al., 2006)
Weevils*	Gamma	Various spp. of weevils	(Lefevre et al., 2004)
Wigglesworthia glossinidia	Gamma	*Glossina* sp.	(Aksoy, 1995)
Asaia sp.	Alpha	*Anopheles stephensi*	(Favia et al., 2007)
Bartonella sp.	Alpha	Various insect species	(Reeves et al., 2005)
Diaphorina citri A*	Alpha	**	(Subandiyah et al., 2000)
Rickettsia sp.	Alpha	Various arthropods	(Sakurai et al., 2005; Perotti et al., 2006)
Wolbachia pipientis	Alpha	Various arthropods and nematodes	(Yen and Barr, 1971; O'Neill et al., 1992)
Diaphorina citri B*	Beta	**	(Subandiyah et al., 2000)
Tremblaya princeps	Beta	Mealybugs	(Thao et al., 2002)
*Adonia variegata**	Bacteroidetes	**	(Hurst et al., 1999)
Blattabacterium cuenoti	Bacteroidetes	Various spp. of cockroaches and termites	(Bandi et al., 1995; Lo et al., 2003)
Cardinium sp.	Bacteroidetes	Various arthropods	(Zchori-Fein and Perlman, 2004; Marzorati et al., 2006)
*Coleomegilla maculata**	Bacteroidetes	**	(Hurst et al., 1997)
*Cryptococcus ulmi**	Bacteroidetes	**	(Gruwell et al., 2007)
Icerya sp.*	Bacteroidetes	**	(Gruwell et al., 2007)
Sulcia muelleri	Bacteroidetes	Various spp. of cicadas, leafhoppers, spittlebugs, treehoppers	(Moran et al., 2005b)

Continued.

Table 1.1 List of Symbiotic Bacteria Included in the Tree in Figure 1.1 (Continued)

Taxon Designation	Classification	Host	References
Uzinura diaspidicola	Bacteroidetes	Various spp. of armored scale insects	(Gruwell et al., 2007)
*Zigia versicolor**	Bacteroidetes	**	(Zchori-Fein and Perlman, 2004)
Spiroplasma sp.	Firmicutes	Various arthropods	(Tsuchida et al., 2006)

Note: P-symbionts are printed in bold. The alpha, beta, and gamma terms stand for individual subdivisions of Proteobacteria. As in Figure 1.1, the taxa are designated by either the names of symbiotic bacteria or those of the host taxa labeled with an asterisk. For the latter case, asterisks are used within the Host column instead of repeating the host name.

P-symbionts and S-symbionts: old terms in a new phylogenetic framework

The "primary" (P) and "secondary" (S) symbionts are among the most firmly established terms in insect symbiosis research. They originated early in the prephylogenetic period of this field (Buchner, 1965) from a highly simplified view: in addition to evolutionary original and nutritionally essential bacteria inhabiting specialized host cells (P-symbionts), some insects carry incidental infections with dispensable or even deleterious bacteria (S-symbionts). Although these terms are frequently used in the literature, their exact meaning remains vague. Based on the above demarcation, several typical traits are usually attributed to these two categories. P-symbionts are large bacteria restricted to specialized host cells (bacteriocytes), inherited exclusively by vertical mode and cospeciating with their host. In contrast, S-symbionts can invade various cells of the host and are often transferred horizontally among unrelated host species or higher taxa.

With the growing diversity of characterized symbionts, this conception is still more difficult to apply and the terms are used with considerable uncertainty. This fact was succinctly stated in the exhaustive review on P-symbionts by Paul Baumann:

> In contrast to this relatively well-defined association between the P-endosymbionts and insects, the association between S-symbionts and insects is currently not amenable to a simple definition, because these organisms form a heterogeneous group with respect to location in the insect and possibly in their function (Baumann, 2005).

For example, when multiple symbiont lineages occur in an individual host, some of the bacteria designated as S-symbionts can display traits very similar to P-symbionts, including congruent phylogeny with the host (Thao et al., 2002). The same uneasiness with this concept has led Takiya et al. (2006) to coin the term *coprimary* symbionts for cases where "two or more symbionts are obligate and ancient bacteriome-associates." Moreover, many phylogenetic analyses demonstrate that there is no fundamental division between these two types of symbionts. As can be seen in Figure 1.1, bacterial lineages display remarkable versatility and can rapidly adapt to various kinds of symbiosis. This dynamic view of symbiont evolution is further supported by additional observations, such as multiple infections and/or symbiont replacements (Lefevre et al., 2004; Takiya et al., 2006; Wu et al., 2006). Experimental evidence of the feasibility of such replacements in highly specialized

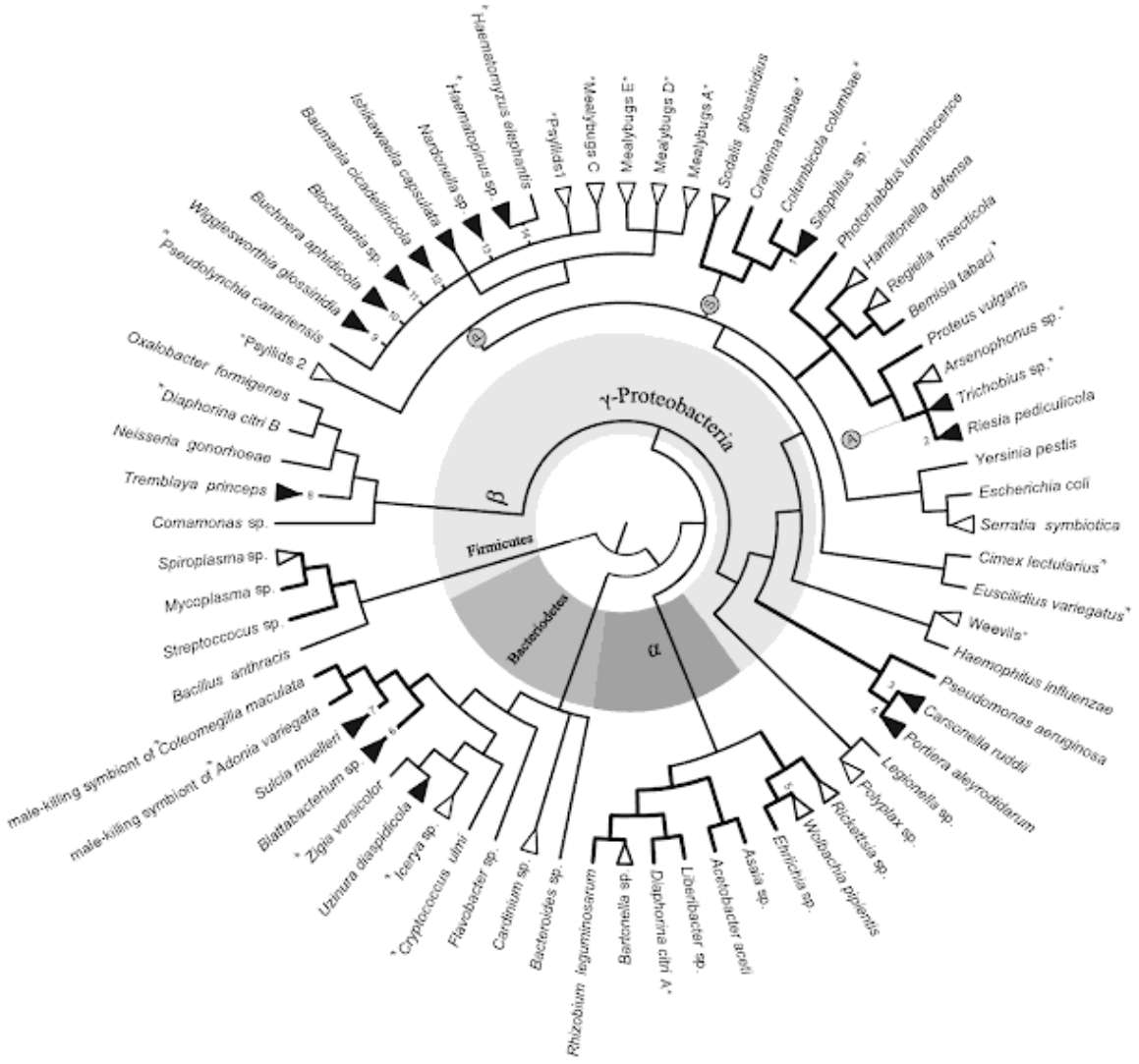


Figure 1.1 Phylogenetic positions of bacteria living in symbiosis with insects. The tree was built using rather subjective criteria (see the text), using published phylogenetic studies. The labels in the tree present either the names of symbiotic bacteria or those of the host taxa (marked with an asterisk). Free-living bacteria are printed in gray. Bold lines designate the relationships for which general consensus exists in the literature. The triangles symbolize monophyletic diversified groups of symbionts described from more than a single host species (solid triangle: lineage phylogenetically congruent with the host). P: putative origin of multiple symbiotic lineages including several of major P-symbionts; S and A designate the nodes discussed in the text. Information on the host taxa and source literature are provided in Table 1.1. Time estimates for the numbered nodes are listed in the Table 1.2.

associations comes from a study demonstrating the capacity of S-symbionts to take over a nutritional role to replace the eliminated P-symbionts (Koga et al., 2003).

As in many other ecologically rooted categories (i.e., parasites vs. commensals, parasite vs. parasitoid), it is unlikely that the diversity of symbiotic bacteria can be fitted into a simple and unequivocal terminological framework. However, to make the following discussions as transparent as possible, we use the P-symbiont and S-symbiont terms in their

most conservative meaning. We set aside the chronological question of the associations' origin and reserve the term P-symbiont for the clearly coevolving mutualists, such as *Buchnera* or *Wigglesworthia*. All other types of symbionts are referred to as S-symbionts, with full awareness of the ecological and functional heterogeneity of this category.

Composition of symbiotic fauna: transitions, losses, and replacements

It has been generally supposed that the compensation for missing nutrients by symbionts is a prerequisite of host survival and adaptive radiation for some feeding strategies. The logical consequence in such cases would be a unique origin and perpetual coevolution between the host and its symbiont. The first phylogenetic descriptions were indeed in accordance with such a model: both aphids and tsetse flies form well-defined groups with specific life strategies and harbor unique monophyletic lineages of symbionts (Munson et al., 1991a; Chen et al., 1999). However, an increasing amount of molecular evidence shows that this strict coevolutionary view has to be relaxed. Multiple infections with distant P-symbionts, losses, and/or replacements often shape insect–bacteria associations. Even the host groups nutritionally dependent on typical P-symbionts can harbor an assemblage of phylogenetically distant symbionts, which have arisen through a series of acquisitions/losses and display a mosaic of different coevolutionary histories. Such a decoupled life strategy and symbiosis history can be demonstrated by the composite P-symbiont fauna of two groups of hematophagous insects, Anoplura and Hippoboscoidea.

As obligatory hematophages, sucking lice (Anoplura) conform to a widely accepted definition of insects dependent on bacterial symbionts (Douglas, 1989) and their nutritional dependence has even been experimentally demonstrated by elimination of symbionts (Aschner, 1934; Puchta, 1955). Because hematophagy is clearly a universal feeding strategy that originated in the common ancestor of extant anoplurans, one should expect to see an uninterrupted chain of louse-symbiont cospeciations. Such a picture is indeed found when louse-symbiont associations are studied at low phylogenetic levels; symbiotic bacteria of the genus *Riesia* have been shown to mirror phylogenetic relationships of their host genera *Pediculus* and *Phthirus* (Allen et al., 2007). However, within the frame of the whole order Anoplura, P-symbionts form an apparently polyphyletic assemblage exceeding even the borders of Enterobacteriaceae (Hypša and Křížek, 2007). Considering their nutritional significance, the polyphyletic nature of louse P-symbionts, contrasting to the perfect coevolution of *Buchnera*-aphid, may seem unexpected. Remarkably, a similar situation is encountered in yet another hematophagous taxon, the dipteran superfamily Hippoboscoidea. Within this group, the tsetse fly–*Wigglesworthia* symbiosis is often mentioned as a typical P-symbiont association analogous to the aphid-*Buchnera* system. However, this view is at least partly due to a long-lasting omission of three "unimportant" groups related to tsetse flies, the families Hippoboscidae, Nycteribiidae, and Streblidae. Only recently it was shown that the families Hippoboscidae and Streblidae carry P-symbiotic bacteria clustering within the genus *Arsenophonus*, at a position very distant from *Wigglesworthia* (Trowbridge et al., 2006).

Both Anoplura and Hippoboscoidea consist of obligatory hematophages and are likely to have differentiated from already hematophagous ancestors. Although the analogy with *Buchnera* would suggest an ancient establishment of symbiosis followed by a long coevolutionary history, both groups seem rather to have undergone multiple symbiont acquisitions, losses, and replacements. Although they provide a good example, sucking lice and hippoboscids are certainly not the only groups that underwent such a rich history of symbiotic associations. Many other insect taxa seem to harbor wealthy assemblages of

symbionts that have arisen from repeated acquisitions and losses (Lefevre et al., 2004; Takiya et al., 2006). As a result of such processes, the gross phylogenetic picture combines several different patterns of symbiotic lineages: (1) long-term coevolution between a P-symbiont and host; (2) multiple infections by different P-symbionts within a host; and (3) replacements of established symbiotic lineages. Each of these specific arrangements represents a different challenge to the phylogenetic reconstruction. For example, most of the typical P-symbiotic lineages can be well recognized as monophyletic groups and their coevolution with the host can be easily established (Munson et al., 1992; Schröder et al., 1996; Chen et al., 1999). The main difficulty rests in determining their position within bacterial phylogeny and is due to far-reaching modifications of their genomes. On the other hand, most S-symbionts retain sufficient phylogenetic information in their sequences to allow for reliable phylogenetic placement. However, it might be difficult to find a scenario reconciling their phylogeny with host distribution. These areas of phylogenetic uncertainty are further discussed in the two following sections.

P-symbionts: a touchstone of molecular phylogenetics

The most interesting and debated node in the whole tree of symbiotic bacteria is undoubtedly the putative origin of many symbiotic lineages within Enterobacteriaceae (Charles et al., 2001). In Figure 1.1, this node is presented in its "maximal" version, encompassing several major P-symbionts and many minor lineages (node P). However, in the published studies, the whole issue has mostly been addressed by analyzing phylogenetic relationships of the two most popular groups, *Buchnera* and *Wigglesworthia*. Although retrieved by a majority of phylogenetic studies, the monophyly of the symbiotic cluster containing these two P-symbiont lineages has been legitimately questioned. This doubt arises because the genomes of P-symbionts meet typical conditions leading to phylogenetic artifacts. Compared to their free-living relatives, P-symbiotic lineages display remarkably high frequency of AT in their sequences. This bias is considered one of the most significant symptoms of genome degradation in symbiotic bacteria. For the first time it has been detected within 16S rDNA and is usually attributed to relaxed selection together with Muller's ratchet occurring in small asexual populations (Moran, 1996; Heddi et al., 1998; Lambert and Moran, 1998; but see Itoh et al., 2002, for alternative explanation). The statistical significance of this compositional shift was later confirmed by Haywood-Farmer and Otto (2003).

Because the difference in nucleotide composition among the lineages seriously violates assumptions implemented in the majority of phylogenetic methods, it results in more or less predictable artifacts. Currently, the most common approach to this problem is an employment of techniques designed to eliminate or at least suppress the effect of compositional heterogeneity. Several such methods have been proposed and this area is undergoing fast advancement. An alternative approach does not rely on extraction of phylogenetic signal by tuning the assumptions to (supposedly) real evolutionary process, but rather to extend the dataset and/or to find alternative sources of information. Genome-wide concatenation of protein-coding genes and extraction of a phylogenetic signal from genome structure are such techniques.

Nonhomogeneous models

Most P-symbiont studies have used phylogenetic methods implementing standard models of molecular evolution, which are based on stochastic processes with two main implicit assumptions: homogeneous base composition and constant substitution rates. Due to

these assumptions, the evolutionary processes can be modeled and analyzed using the time-reversible Markov chain model as the methodological basis. Consequently, if any force directs substitution processes, the assumption of time-reversibility becomes violated. A typical, well-known example of such selection-driven change is the compositional difference between 16S rRNA genes of thermophilic and mesophilic bacteria. When analyzed in the context of thermophils, the mesophilic bacteria *Deinoccus* and *Bacillus* cluster as sister groups, in contrast to strong evidence for their polyphyly. This conflict has been repeatedly attributed to convergent selection-driven evolution of thermophils toward a GC-rich genome (Mooers and Holmes, 2000; Foster, 2004). A similar effect can be seen in the AT-rich sequences of symbiotic bacteria. Particularly in 16S rDNA analysis, this phenomenon can play a crucial role, because long stretches within the transcribed rRNA loops can accommodate an enrichment of AT residues.

This problem does not have any simple solution. Initial attempts to cope with it relied on distance calculations eliminating the effect of composition heterogeneity, particularly the paralinear (Logdet) method (Lake, 1994; Lockhart et al., 1994) or alternative distance formula suggested by Galtier and Gouy (1995). However, the distance methods are generally considered an inferior phylogenetic tool compared to the maximum parsimony (MP), maximum likelihood (ML), or Bayesian analysis. It is therefore understandable that the nonhomogeneous approach was soon introduced into the maximum likelihood framework.

The model developed by Yang and Roberts (1995) extended the well-known HKY85 (Hasegawa et al., 1985) substitution model by introducing different compositional parameters for each tree branch. Although this algorithm is in principle capable of dealing with nonhomogeneous sequences, this model is too parameter rich and thus computationally demanding. Moreover, the necessity to estimate parameters from the data is a potential source of topological distortions. To overcome these difficulties, Galtier and Gouy (1998) simplified the model by replacing the HKY basis with T92 (a single parameter for G + C) (Tamura, 1992). It was only this new version of the nonhomogeneous model that was subsequently used to test the monophyly/polyphyly of the P-symbiotic lineages (Herbeck et al., 2005). This study brought the first strong evidence favoring P-symbionts polyphyly. However, it has not settled the issue at all. On the contrary, several authors expressed their dissent with the polyphyletic view and tried to prove the opposite.

The main problem is that while there is no doubt about the superior performance of nonhomogeneous model(s) in some particular cases, it may be extremely difficult to predict their behavior for various matrices and datasets. Indeed, selection of a proper model is one of the very central issues of ML methodology. A well known property of evolutionary models is that their predictive power decreases with additional parameters (Posada and Buckley, 2004; Steel, 2005). The nonhomogeneous model applied to the P-symbiont phylogeny uses a free compositional parameter(s) on each branch, which may rapidly lead to the over-parameterization of the analysis with the increasing number of branches in the tree. Ultimately, this property is a reason why the nonhomogeneous technique may not be particularly suitable for solving the P-symbionts issue. To decrease the complexity of the nonhomogeneous model, Foster (2004) suggested that instead of introducing many free parameters along a tree, an application of only a few vectors of composition is sufficient to handle compositional changes. He used the aforementioned thermophile problem to test this method and showed that it can indeed be solved by introducing only two vectors. To find the optimal solution, he employed Bayesian analysis to test the fit of the nonhomogeneous model to data. Lately, the nonhomogeneous models are being further developed in several different directions (Blanquart and Lartillot, 2006; Gowri-Shankar and Rattray,

2007). None of these new Bayesian-based methods, optimizing the number of parameters, have so far been used to address the P-symbiont issues. However, rapid development of the techniques extracting phylogenetic signal from heterogeneous sequences indicates that it would be premature to draw any conclusion on P-symbionts monophyly/polyphyly from the analyses that have been reported.

Multigene approach

Leaving aside the question of a "proper" algorithm, it should be admitted that both monophyletic and polyphyletic arrangements are only poorly supported by the 16S rRNA-derived signal. In other words, the 16S rRNA gene, the most frequently used marker in bacterial phylogeny, is incapable of solving the relationships among P-symbionts. This observation is not surprising because such insufficiency of rRNA genes is frequently observed at various phylogenetic levels in many groups of organisms. Various protein-coding genes have been used as an alternative source of phylogenetic information (Degnan et al., 2004; Casiraghi et al., 2005; Moran et al., 2005a; Baldo et al., 2006; Fukatsu et al., 2007). Although they can yield better phylogenetic resolution at some particular nodes in dependence of their evolutionary tempo, they do not provide any fundamental advantage if used in single-gene matrices.

The only way of overcoming the lack of reliable information is an extension of the dataset with additional sequences and employment of the multigene approach. Although seemingly simple and straightforward, the method of adding new genes is not free of potential troubles. The typical bacterial genome is a flexible assemblage of genes undergoing frequent structural changes (Snel et al., 2005). Some of these processes may hinder selection of suitable universal markers among hundreds of possible candidates. For example, loss of genes leads to absence of a given phylogenetic marker in some bacterial lineages. This situation may be particularly frequent in symbiotic bacteria that undergo rapid and dramatic loss of many genes; substantial reduction of genome size can be seen in all of the completely sequenced genomes of P-symbionts (Shigenobu et al., 2000; Akman et al., 2002; Gil et al., 2003; Nakabachi et al., 2006; Wu et al., 2006) and has been observed even in the presumably young symbiotic lineage *Sodalis glossinidius* (Toh et al., 2006). Moreover, different nutritional constraints in various host–symbiont associations lead to differential preservation/loss of various sets of genes in different symbionts. Thus, on their hypothetical pathways from free-living bacterium to highly specialized symbionts, *Buchnera* and *Wigglesworthia* reduced their genomes to approximately 583 and 621 coding genes, respectively (Shigenobu et al., 2000; Akman et al., 2002); only 69% of these genes are shared by both lineages (Akman et al., 2002). Similar functional complementarity between two different symbiotic genomes, although not based on complete genome sequences, was recently reported for the genera *Sulcia* and *Baumannia* (Wu et al., 2006). If such small genomes are to be analyzed together with free-living bacteria, genes have to be identified that are present in all of the included genomes. To make the situation even more complicated, successful identification of homologous genes is only one prerequisite, but does not itself guarantee a consistency of phylogenetic signal. At least two additional processes may disturb phylogenetic reconstruction. Duplications are a known and much feared source of paralogs, which are further inherited during the speciation process. A random sampling of paralogs from different lineages during the phylogenetic analysis can be a source of serious topological inaccuracies. Finally, even worse phylogenetic inconsistencies may arise due to horizontal gene transfer (HGT), a process that introduces phylogenetically distant xenologs into bacterial genomes.

The significance of duplication and HGT for phylogenetic inference in bacteria has not been fully elucidated. Generally, it is supposed that duplications in prokaryotes are less deleterious than in eukaryotic organisms. By contrast, the HGT is often detected in bacteria and has sometimes even been considered as one of the main forces shaping bacterial genomes. However, current views on this issue are largely dependent on the methods used to estimate overall HGT frequency (Lerat et al., 2003; Susko et al., 2006; Doolittle and Bapteste, 2007). For example, a conservative view, with vertical inheritance playing a predominant role in the bacterial genome structure, has been voiced by Lerat et al. (2003). These authors assessed the overall compatibility of individual single-gene matrices with selected topologies. To achieve this, they postulated phylogenetic congruence as a null hypothesis and used the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999) to identify an HGT by its rejection. Comparing 13 genomes of γ-Proteobacteria, they showed that the universally present orthologs suffer only negligible frequency of HGT: of 205 genes included in the analysis, 203 produced mutually compatible topologies. When used for phylogenetic inference within a concatenate matrix, this respectable set of genes produced a monophyletic and well supported branch of *Buchnera* + *Wigglesworthia* that was preserved even after removal of the AT-rich codons. On the other hand, such a low level of HGT has recently been questioned by Susko et al. (2006). They adopted methods from functional genomics to visualize the congruency within the core gene set suggested by Lerat et al. (2003) and concluded that around 10% of the genes may have resulted from HGT. In their discussion, they further postulate that using congruence as a null hypothesis and searching for the significant incongruence necessarily leads to underestimation of the HGT level. A similar opinion about the considerable occurrence of HGT has also been reached by Doolittle and Bapteste (2007) using an entirely different source of evidence than phylogenetics. Considering the whole spectrum of HGT frequency estimates, stretching from almost zero (Ge et al., 2005) to more then 60% (Lerat et al., 2005; Dagan and Martin, 2007), it is hard to assess the possible effect of this phenomenon on the selection of a suitable set of phylogenetic markers.

Genome structure

Changes in nucleotide sequences are not the only trace evolutionary processes left on the genome. Various other characters such as gene fusion (Philippe et al., 2000; Stechmann and Cavalier-Smith, 2002), codon reassignment (Castresana et al., 1998), RNA secondary structure (Billoud et al., 2000; Swain and Taylor, 2003), or gene arrangement (Lavrov et al., 2004) originate by unique evolutionary events and may retain valuable phylogenetic information (Rokas and Holland, 2000). Although some of these "idiosyncratic" markers (Murrell et al., 2003) may be affected by the same processes as nucleotide sequence (e.g., codon bias, secondary structure), others are likely to be sequence-independent. The most popular source of such markers by far is the arrangement of genes within a genome.

Although this tool has only recently been "discovered" for bacterial phylogeny and used in a few studies, the idea itself is relatively old. The reason for its omission in bacterial phylogeny rests mainly in the computation demands connected to genome-wide analysis. Most studies relying on this method have focused on eukaryotes and dealt with mitochondrial genomes (Lavrov et al., 2004; Negrisolo et al., 2004; Larget et al., 2005; Podsiadlowski et al., 2007). Their results showed that, like other sources of phylogenetic data, gene order can be helpful to solve some particular problems (Lavrov et al., 2004), whereas it fails in others (Negrisolo et al., 2004). Among the few analyses performed on prokaryotes, some are of very low informative value with respect to P-symbiont issues, because they

did not include P-symbiotic bacteria at all (Comas et al., 2006) or contained only a single P-symbiotic lineage (Korbel et al., 2002). The only analysis that explicitly addressed the monophyly/polyphyly of P-symbionts introduced yet another solution into the repertoire of suggested scenarios by arranging *Wigglesworthia* and *Buchnera* in a paraphyletic manner with respect to several free-living bacteria (Belda et al., 2005).

Evolutionary implications of P-symbionts monophyly/polyphyly

With our current state of knowledge, the answer to the P-symbionts issue can only rest on personal opinion. Seen from this perspective, the complex analysis of Comas et al. (2007) appears to provide the most convincing evidence on the monophyly of at least *Blochmannia* + *Wigglesworthia* + *Buchnera*. Although the opposite conclusion was reached in two recent studies based on different methods (Belda et al., 2005; Herbeck et al., 2005), their results are not mutually compatible. Under the nonhomogeneous model, *Buchnera* and *Wigglesworthia* were placed at very distant phylogenetic positions (Herbeck et al., 2005), whereas in gene-order analysis they branched as closely related lineages in paraphyletic arrangement (Belda et al., 2005). Considering these topological differences and the highly experimental nature of the utilized methods, the results indicating the non-monophyly of P-symbionts should for the present be treated with caution.

However, regardless the methodological reliability of various published analyses, at least two circumstances make their evolutionary interpretation uncertain. First, these studies only used datasets from a few selected symbiotic lineages. Because, as a rule, the accuracy of phylogenetic inference depends on dataset completeness, an inclusion of additional symbionts might affect both topology and support of the tree. In Figure 1.1, we show all symbiotic lineages that were placed by some analysis into the P-symbiont cluster. To the best of our knowledge, no published study included all, or at least a majority, of these symbionts. In the current literature, there is an obvious bias toward the most popular symbionts, namely *Buchnera* and *Wigglesworthia*, while others are often neglected. At least with genome-wide studies this situation can be attributed to the limited availability of complete genomes. Second, and more important, even the final solution of the monophyly/polyphyly problem does not provide a definitive answer to the origin of P-symbiotic bacteria. Upon retrieving the monophyletic *Buchnera-Wigglesworthia* clade, Canbäck et al. (2004) suggested a common origin of these two lineages from a symbiotic ancestor. To support this view, they referred to similar results obtained by Lerat et al. (2003) in their broad phylogenomic analysis. However, no such strong statement was made in the latter study. On the contrary, the authors rightly stated that due to the lack of any gene common for *Buchnera* + *Wigglesworthia* but absent in other related bacteria, the issue of symbiosis origin cannot be effectively addressed.

Obviously, the monophyly of the P-symbiotic lineage does not necessarily imply a unique transition from a free-living bacterium. Considering the evolutionary distance among homopterans and dipterans, such a hypothesis would require an establishment of symbiosis extremely deep within the insect phylogeny, followed by a high frequency of secondary losses. This is not entirely inconceivable because a similar process has already been postulated for another insect–symbiont association. A newly described bacterium *Sulcia muelleri* displays "patchy" distribution across several groups of Auchenorrhyncha (Moran et al., 2005b). Considering the significant congruence between *Sulcia* and host phylogenies, this distribution has been attributed to a long-term cospeciation accompanied by many losses. However, even this impressive story is incomparable to that implied by *Buchnera-Wigglesworthia* common ancestry (Takiya et al., 2006). Current estimates of

insect groups' divergence indicate that such bacterium would have to maintain its symbiosis for the period of time reaching almost 400 My (Gaunt and Miles, 2002). This view is undoubtedly much less parsimonious than supposing an independent origin of several P-symbionts within a cluster of related lineages.

Coevolutionary reconstructions in symbiotic associations

Tree congruence

Coevolutionary reconstructions are among the most precarious enterprises in phylogenetics. Their main aims are twofold: (1) to identify whether phylogenetic congruence between two associated organisms is higher than expected by chance, and (2) to localize and explain the observed incongruences. One of the biggest problems in coevolutionary analysis is that incongruences can arise from fundamentally different sources: real evolutionary events (e.g., host switches and/or multiple colonizations) or methodical artifacts (incorrect topology, lineage duplication). During the several decades of theoretical work, mostly devoted to parasitological research, many algorithms were developed to disentangle these processes, each of them based on different evolutionary assumptions (Ronquist, 1995; Huelsenbeck et al., 2000; Johnson et al., 2001). In insect–symbiont research, the coevolutionary framework takes a specific form. For the clearly mutualistic P-symbionts, an entire phylogenetic congruence due to exclusively vertical transmissions is implicitly presumed. In contrast, frequent horizontal transfers seen in many "guest microorganisms" (Douglas, 1989) suggest random distribution and usually do not provoke any rigorous coevolutionary approach.

Perhaps due to these circumstances, insect–symbiont associations were only rarely addressed by rigorous coevolutionary methods. In simple cases, where both coevolving counterparts carry a strong and unequivocal phylogenetic signal, congruency can be easily determined by looking at the trees. Indeed the first confirmation of the concept of strict coevolution between insects and typical P-symbionts relied on such a visual assessment (Munson et al., 1991a; Schröder et al., 1996; Chen et al., 1999). However, many recent findings suggest that insect–symbiont associations should be seen as more dynamic systems built by multiple acquisitions, losses, replacements, and horizontal transfers. At the same time, because phylogenetic inference is far from being impervious to topological artifacts, an increasing number of taxa leads inevitably to phylogenetic inaccuracies. These, in turn, introduce false incongruence even into entirely coevolving associations.

The fact that different datasets can yield different topologies, even if produced by the same phylogeny, is common knowledge in phylogenetic theory (Hipp et al., 2004). Several cases of partial incongruences observed between insects and their symbionts during the last few years (Thao et al., 2000b; Downie and Gullan, 2005; Takiya et al., 2006; Gruwell et al., 2007) show that this danger is not only theoretical but may blur many strictly coevolving associations. In all of these studies, the authors identified significantly non-random association between insects and their symbionts, but always accompanied by several incongruent relationships. As a rule, the authors tend to explain such incongruences as being caused by a weak phylogenetic signal rather than by real biological events (e.g., horizontal transfers). This attitude should not be seen as mere prejudice: there are certainly sound reasons to take the horizontal transfers as unlikely events. The intimacy of these associations leads to the establishment of various precisely tuned adaptations that might considerably limit or entirely block any horizontal transfer. On the other hand, the degree of observed incongruence is sometimes relatively high and should not be easily dismissed

without proper analysis. For example, after finding 75% of nodes congruent between mealybugs and their P-symbionts, Downie and Gullan (2005) explained the incongruences mostly as phylogenetic artifacts, but they also acknowledged that rare horizontal transfers may have taken part in the overall pattern.

Generally, the coevolutionary approach rests on two different methodological bases, the event- and the data-based methods. The event-based analyses, such as Treefitter or TreeMap (Ronquist, 1995; Charleston and Page, 2002), try to fit two incongruent phylogenies into the "best" coevolutionary hypothesis. To do this, they optimize various evolutionary events (e.g., cospeciations, duplications, switches, and losses) on already reconstructed incongruent trees. In other words, the conflicts between the two topologies are readily accepted as real phylogenetic incongruences and the program attempts to fit the best biological explanation onto the trees. This approach is particularly reasonable in a parasitological context, where some degree of incongruence is expected as a rule (Clayton et al., 2004). In symbiosis, this kind of reconstruction has only rarely been used. For example, Takiya et al. (2006) tested the topologies of Cicadellidae and their P-symbionts with Treefitter and concluded that despite observed differences, the trees display significant overall congruency that cannot be explained by mere chance. However, in contrast to host–parasite associations, whenever coevolutionary signal is detected between insects and their symbionts, a strict congruence is the most common expectation. Thus, analysis is mainly centered on data-based methods that allow testing whether such null hypotheses can be rejected on the basis of available data, rather than the actual reconstruction of a coevolutionary scenario.

Compatibility of host and symbiont data

The question of whether two or more partitions of the dataset come from the same phylogeny is very common in phylogenetics. Incongruence length difference (ILD) (Farris et al., 1995a; 1995b) and its variants are the most frequent tests used to address such problems. Although in most cases these partitions are represented by sets of different genes collected from the same taxa, this method can be easily adopted to matrices combined from two different organisms. Due to this property, the ILD test can be applied in coevolutionary studies (Johnson et al., 2001; Lopez-Vaamonde et al., 2001; Hughes et al., 2007).

For insect symbiont associations, the ILD test was used in a few studies and allowed for both the rejection and corroboration of a null hypothesis. Results from analysis on psyllids and *Carsonella* (Thao et al., 2000b) as well as Cicadellidae vs. *Sulcia* (Takiya et al., 2006), showed that despite topological incongruences, the null hypothesis of strict cospeciation could not be rejected when the data matrices were taken as the base. A different case was reported by Downie and Gullan (2005); for a 75% topological correspondence between mealybugs and their symbiont, the overall congruence of the datasets was rejected by ILD. Considering the similarities of all of these insect–symbiont associations and the different implications of the ILD tests, knowledge of the latter's properties and reliability becomes particularly important. There have been many debates in the phylogenetic literature on what kind of information and sources of error should be expected from this test (Yoder et al., 2001; Hipp et al., 2004; Ramirez, 2006; De Vienne et al., 2007; Quicke et al., 2007). With respect to the insect–symbiont studies, the most disquieting weakness of ILD is its sensitivity to unequally distributed homoplasies among data partitions (Dolphin et al., 2000). In a coevolutionary framework, such difference in noise content between the two partitions is very likely. Most recently, a new modification of the test, arcsine ILD, has been suggested to suppress this tendency (Quicke et al., 2007). At the time of this chapter preparation,

no study on insect–symbiont association had applied this new approach, and its power remains to be tested.

Another problem with the ILD approach is that while assessing overall congruency, it is not able to identify taxa responsible for the incongruences. A simple remedy for this shortcoming was proposed by Johnson et al. (2001). Their approach consists of removing taxa (e.g., host–parasite pairs) from the dataset until significant congruence between the partitions is achieved. This restricted set of taxa is then used to build a phylogenetic backbone common for both counterparts. Because the removed taxa are supposed to be incongruent, their position is inferred by a subsequent phylogenetic analysis constrained with the backbone topology, and the resulting trees are finally compared by a tree-based method. The efficiency of this technique to localize the source of incongruence has been recently demonstrated on a host–parasite system (Hughes et al., 2007). Due to its capacity to identify a maximal set of entirely congruent taxa, this approach might be particularly valuable for investigating insect–symbiont associations. To the best of our knowledge, it has not yet been applied to any of the known insect symbionts.

A different approach in the identification of incongruent taxa is offered by the method of partitioned Bremer support (PBS) (Baker and DeSalle, 1997). Dividing the overall Bremer support of each node to individual partitions, this method can highlight the taxa minimally contributing or even contradicting the particular node. In insect–bacteria coevolution, the partitioned Bremer has recently been used to identify possible incongruences between mealybugs and their insect hosts (Downie and Gullan, 2005). As agreed by many authors, none of the incongruence testing methods can provide definitive proof for tree congruence/incongruence and should be combined with other sources of evidence.

Tree confidence

Considering the almost inevitable errors in phylogenetic inference, the "best" tree (e.g., ML tree) is not necessarily the true tree. Thus, even if the host and symbiont trees differ from each other, the symbiont tree may not be a significantly worse explanation of the host data, and vice versa. In the realm of ML, a set of methods has traditionally been available to test the significance of tree differences with respect to the underlying data. The most popular tests are the Kishino-Hasegawa test (KH test) (Kishino and Hasegawa, 1989) and its later modifications. Although the likelihood-based tests are extremely popular and were often used, their application is not always straightforward. In the analyses of symbiotic associations, the competing topologies are trees derived directly from the data. It has been convincingly shown that under such circumstances, the KH test is not applicable (Shimodaira and Hasegawa, 1999; Goldman et al., 2000b). A modified version, the SH test developed by Shimodaira and Hasegawa (1999), overcomes this obstacle, but pays a considerable cost in the form of increased conservativeness (Goldman et al., 2000a).

For insect–symbiont coevolution, this kind of test was used, for example, by Clark et al. (2000) to assess the congruence between *Buchnera* and aphids of the genus *Uroleucon*. An important outcome of this study is the finding that an ML-based test can reject compatibility of the host and symbiont data even under high topological congruency. The author concluded that in the observed aphid-*Buchnera* system, this might be due to the heterogeneity of evolutionary process in different regions of the aphid mitochondrial DNA. Since this report, several other tests were proposed, most of them based on the ML method. Shi et al. (2005) tested several of these approaches in simulation studies and stressed the good performance of the SOWH test, a method based on parametric bootstrap (Goldman et al., 2000a). There is currently only limited experience on performance of this and other tests in

the coevolutionary field. To the best of our knowledge, the only application of the SOWH test to insect–symbiont coevolution was done in the study of Downie and Gullan (2005). Comparing phylogenies of mealybugs and their P-symbionts, the authors were able to show that the incongruence between the two topologies is not caused by sampling error.

Molecular clock and calibration of evolutionary rates

An extremely appealing method of assessing a coevolutionary scenario when topology-derived evidence is not conclusive is the comparison of speciation times. Such an approach is well known in the coevolutionary field and has even been implemented into some of its programs (Merkle and Middendorf, 2005; Charleston and Page, 2002). Ultimately, the dating of evolutionary events relies on calibration with fossil records. However, for many organisms, including parasites and symbiotic bacteria, no fossil records are usually available. The possible solution to this problem was proposed via the molecular clock concept (Avise, 1994). Where the cophylogenetic scenario is strongly supported by topological correspondence, fossil-based times of hosts can, in principle, be used to subsequently calibrate the evolutionary rate of its parasite/symbiont. For insect symbionts, this approach has been applied to several groups, such as aphids and *Buchnera* (Moran et al., 1993), cockroaches and their symbionts (Bandi et al., 1995), or psyllids and *Carsonella* (Thao et al., 2000b). The results of such calibrations revealed that the evolutionary tempo in symbiotic bacteria is considerably higher than the rates calculated for free-living bacteria (Moran et al., 1993; Moran, 1996). Such a result is not surprising as symbiotic bacteria are known to more quickly accumulate mutations due to regular and severe bottlenecks and relaxed selection.

Hypothetically, the calibrated evolutionary rates of these symbiotic bacteria could be used to determine the divergence times of other symbionts and decide between coevolutionary and multiple-origin scenarios. This method was applied to related symbionts from aphids and whiteflies (Darby et al., 2001) or anoplurans and rhynchophthirines (Hypša and Křížek, 2007). Unfortunately, despite its undeniable attractiveness, the calibration of evolutionary tempo seems to be of very limited applicability in insect–symbiont investigations. Calculations performed for various symbiotic lineages brought an unpleasantly broad range of time estimates (Table 1.2). Most important from this point of view is the considerable divergence in the evolutionary rates between two symbionts coevolving with the same host, *Sulcia* and *Baumannia* (Takiya et al., 2006).

Specificity without coevolution

An interesting phenomenon that has been largely overlooked in the insect–symbiont field is the occurrence of a false coevolutionary signal, i.e., a signal created in absence of actual cospeciation events. In such cases, the probability of a successful host switch may be constrained by phylogenetic relatedness of the hosts. As a result, the symbiont can be restricted to particular host taxa due to functional constraints rather than common evolutionary history. Empirical examples of such false congruence due to preferential switching come from the studies on primate lentiviruses (Charleston and Robertson, 2002) or host-plant specificity in psyllids (Percy et al., 2004). From a theoretical point of view, the conditions favoring false congruence were recently analyzed by De Vienne et al. (2007). They demonstrated that the higher probability of a parasite colonizing a species related to the current host can indeed lead to an extremely high degree of congruence. It would be

Table 1.2 Divergence Times Estimated for the Nodes Depicted in Figure 1.1

Node	Symbiotic Bacterium	Divergence Time (MYA)	References
1	*Sitophilus* endosymbiont	50–100	(Heddi et al., 1998)
		100	(Dale et al., 2002)
		25	(Lefevre et al., 2004)
2	*Riesia*	5,6	(Allen et al., 2007)
3	*Carsonella*	100–250	(Thao et al., 2000b)
4	*Portiera*	100–200	(Thao and Baumann, 2004b)
5	*Wolbachia*	58–67	(Werren, 1997)
		60–100	(Duron and Gavotte, 2007)
6	*Blattabacterium*	135–300	(Moran and Wernegreen, 2000)
		140–145	(Lo et al., 2003)
7	*Sulcia*	260–280	(Moran et al., 2005b)
8	*Tremblaya*	100–200	(Thao et al., 2002)
9	*Wigglesworthia*	40	(Moran and Wernegreen, 2000)
10	*Buchnera*	160–280	(Moran et al., 1993)
		100–200	(Clark et al., 2000)
		150–250	(Moran and Wernegreen, 2000)
11	*Blochmannia*	50–100	(Moran and Wernegreen, 2000)
		90–110	(Degnan et al., 2004)
12	*Baumannia*	80–175	(Moran et al., 2003)
		70–100	(Moran et al., 2005b)
		138–475	(Takiya et al., 2006)
13	*Nardonella*	100	(Lefevre et al., 2004)
14	*Haematopinus* endosymbiont	190–230	(Hypša and Křížek, 2007)

interesting to consider whether some of the insect–bacteria associations might have arisen from preferential switching instead of cospeciation processes.

Currently, there is no direct evidence of a similar host-mediated constraint affecting the distribution of bacterial symbionts. In the frequently switching bacterium of the genus *Wolbachia*, most transfers take place between unrelated hosts, and some can bridge as large phylogenetic spans as between different arthropod groups or even arthropods and nematodes (Sironi et al., 1995). However, in several cases, the monophyletic cluster of *Wolbachia* was retrieved from closely related hosts. Most recently, such a phylogenetically tight group was found in several scorpion species of the genus *Opisthophtalmus* (Baldo et al., 2007). Nevertheless, due to the lack of reliable host phylogeny, the decision between a coevolutionary and switching scenario cannot be easily made and the evolutionary interpretation remains unclear. The situation is even more conspicuous in another group of typical S-symbionts: the monophyletic lineage of *Sodalis glossinidius* (Dale and Maudlin, 1999) has been detected exclusively from *Glossina* species. At the same time, the extremely low molecular diversity among *S. glossinidius* isolates strongly indicates a recent independent origin of *S. glossin-*

idius infections in individual tsetse species. This observation led Aksoy et al. (1997b) to explain the *S. glossinidius* restriction to a narrow host group by functional constraint. If this explanation is correct, the *S. glossinidius–Glossina* association could provide an example of preferential switching among symbiotic bacteria.

It should also be stressed that the view of the recent establishment of *S. glossinidius* in tsetse flies rests mainly on comparison with another symbiont, *Wigglesworthia glossinidia*. This symbiont obviously underwent a long cospeciation history with tsetse flies, resulting in congruent phylogeny with *Glossina* and the considerable genetic diversification of the *Wigglesworthia* extant lineages (Chen et al., 1999). Compared to 2%–6% 16S rDNA divergence among the *Wigglesworthia* lineages, the near identity of *S. glossinidius* 16S rRNA genes seems to support its recent origin (Aksoy et al., 1997a). This view is certainly plausible and a likely explanation of *S. glossinidius* evolution. However, as noted in the previous section, the dramatic difference in evolutionary tempo observed in the two different symbionts from a single host is a warning that maximal caution should be taken when using sequence divergence as a sole criterion for dating the age of symbiosis.

The nature of *Sodalis–Glossina* association has been discussed in several studies with contradictory evidence. For example, the finding of practically identical isolates in various tsetse species, detection of several genetically distinct strains in a single species (Geiger et al., 2005; Geiger et al., 2007), and successful interspecific transfer of *S. glossinidus* among glossinas (Weiss et al., 2006) point toward a recent origin. In contrast, the considerable degradation of the *S. glossinidius* genome (Toh et al., 2006) is a typical sign of long-term symbiosis rather than recently established lineages. Thus, distribution and phylogeny of *S glossinidius* cannot be unequivocally explained. Moreover, a similar pattern can be seen in yet another lineage, the bacteria of the genus *Arsenophonus* (Gherna et al., 1991). Because both *Sodalis* and *Arsenophonus* represent lineages with a rapidly increasing number of closely related symbionts reported from phylogenetically distant hosts, their investigation poses methodological problems fundamentally different from those connected to P-symbionts.

Low-level phylogeny

An attempt to reconcile phylogeny and host distribution within a consistent evolutionary framework becomes particularly difficult when the analyzed group contains both types of symbionts, coevolving mutualists, and "accidental" temporary infections. Moreover, at low phylogenetic levels, inference is further complicated by additional specific obstacles. The first is that during a short evolutionary time, the sequences gather only a limited amount of phylogenetic information. As an example, in their coevolutionary analysis on Diaspidae and *Uzinura* symbiont, Gruwell et al. (2007) reported a majority of incongruent nodes being located close to the tree terminals. They concluded that this feature is most likely due to a scarcity of information at the "shallow phylogenetic level." Because most of the S-symbiotic associations are supposed to be of relatively young origin, their analyses are likely to suffer this shortage of information. This problem is notoriously evident in *Wolbachia*, where the exclusive usage of 16S rRNA has been largely abandoned and most analyses are now based on several rapidly evolving genes (Zhou et al., 1998; Bordenstein and Rosengaus, 2005; Casiraghi et al., 2005; Vaishampayan et al., 2007) and/or MLST-based approaches (Baldo et al., 2006; Paraskevopoulos et al., 2006). A second typical problem is that when working with closely related taxa, several genetic factors, not encountered at higher phylogenetic levels, can entirely mislead inference of phylogeny and blur a coevolutionary pattern. Various forms of polymorphism followed by incomplete lineage sorting, or intragenomic heterogeneity belong to the best known and often studied phenomena.

Intragenomic heterogeneity

In bacteria, the intragenomic heterogeneity of the 16S-23S-5S rRNA operon (rRNA operon thereafter) and its influence on phylogenetic reconstruction in closely related species has previously been shown in a taxonomically broad range of free-living species (Luz et al., 1998; Pettersson et al., 1998; Yap et al., 1999; Marchandin et al., 2003; Boucher et al., 2004; Lin et al., 2004). Moreover, it was studied in detail for the *Escherichia/Shigella* branch (Cilia et al., 1996; Garcia-Martinez et al., 1996; Anton et al., 1998, 1999; Martinez-Murcia et al., 1999), where the presence of two basic types defined according to specific tRNA gene within the intergenic spacer region (ISR) was demonstrated (Condon et al., 1995). In their comparative study of 55 bacterial spp., Coenye and Vandamme (2003) showed that the intragenomic heterogeneity can reach up to 19 nucleotide differences.

Despite all this evidence, the potential significance of this phenomenon for reconstructing coevolution in insect–symbiont associations has not yet been noted. An obvious reason for this omission is that most of the detailed coevolutionary studies have been devoted to *Buchnera*, *Wigglesworthia*, and other P-symbionts (Aksoy et al., 1997b; Chen et al., 1999; Clark et al., 2000; Sauer et al., 2000; Thao et al., 2000b). In these bacteria, the sequencing of 16S rDNA has proven to be an effective tool for deriving clear coevolutionary patterns. Thus, the intragenomic heterogeneity may not seem to pose any problems to phylogenetic analyses in symbiotic bacteria. It is, however, important to realize that the P-symbionts possess two unique features, both of which result from genome degradation and economization. First, only a single copy of the rRNA operon is present in their genomes. Second, these associations are old and the evolutionary rates considerably exceed those calculated for free-living bacteria (Moran, 1996); consequently the branches in the tree are usually long and well differentiated. In non-bacteriocyte symbionts, the situation is completely different, resembling that of free-living bacteria. For example, despite the considerable degradation of its genome, *S. glossinidius* still carries seven copies of the rRNA operon (see the complete genome Acc. No. NC_007712). For *Arsenophonus*, the complete genome is not yet available, but several lines of indirect evidence suggest that it also carries more than a single copy of the rRNA operon. First, the presence of at least two copies was reported by Thao and Baumann (2004a) in their coevolutionary study on *Arsenophonus* and whiteflies. They detected two different types of operons that were distinguishable according to tRNA genes located within the ISR. To avoid misleading phylogenetic information, they removed the ISR sequences from the alignments and used only the rRNA genes in their analysis. Based on these sequences, they reported a phylogenetically independent distribution of *Arsenophonus* bacteria within their hosts, and attributed this distribution to multiple acquisition of the symbiont. Second, related bacteria for which the whole genome is available (e.g., *Photorhabdus luminescens*; Acc. No. NC_005126) typically carry seven copies of this operon, which is consistent with many other γ-Proteobacteria.

An interesting question with strong implications for coevolutionary analyses is the degree of divergence among individual copies. This parameter determines how deep into the tree the influence of heterogeneity can reach. When individual copies of rRNA operon are extracted from the whole genomes available in the Genbank, they display a remarkable variance of intragenomic heterogeneity across the taxa. For example, very low variability is seen within *Salmonella* (Acc. No. NC 004631) or *Haemophilus* (Acc. No. NC 000907). On the other hand, in *Escherichia/Shigella*, the effect of 16S rRNA heterogeneity reaches as deep as the divergence point of the two genera (Cilia et al., 1996). This split has been estimated to have happened approximately 60 to 180 Mya (Ochman and Wilson, 1987). It is interesting to see that such a time span exceeds divergence times of many insect species and higher

taxonomic groups (Gaunt and Miles, 2002). It thus remains to be seen whether the diversity among rRNA operons in *Arsenophonus* and similar bacteria can affect our current view on their distribution in various host species.

Diversity and sample

In addition to purely methodological issues connected to phylogenetic inference, coevolutionary reconstructions are burdened by another problem: the reliability of any evolutionary interpretation depends on the sufficiency of the analyzed sample. The current pace of new symbiont descriptions being published and sequences being deposited into databases indicates that our knowledge on diversity as a whole is still incomplete. Various screenings focused on a taxonomically restricted group of insects have resulted in identification of a wealth of associated bacteria, which have been overlooked for a long time (Reed and Hafner, 2002; Weinert et al., 2007). The impact of such inadequate knowledge on overall diversity and distribution is particularly felt when analyzing frequently switching bacteria, for which coincidental synchronic detection in several unrelated host taxa is unlikely.

The problem can be illustrated by the history of *Arsenophonus* and *Sodalis* groups. Upon their description, both of these bacteria were known as unique lineages phylogenetically isolated from other symbiotic taxa, each of which was reported from a single host group (*Glossina* for *Sodalis*, *Nasonia* for *Arsenophonus*). This idea lasted until the description of additional members of both lineages (Figure 1.2B). In *Sodalis*, it took almost a decade since its phylogenetic characterization (Aksoy et al., 1995) to attain the description of the first related lineage (Lefevre et al., 2004). Moreover, the new *Sodalis* lineage turned out to be a well-established P-symbiont coevolving with its host group, the weevils of the genus *Sitophilus*. The picture became even more complex after recent descriptions of two additional lineages. First, a sister lineage to the *Sitophilus*-derived symbionts was found in a bloodsucking hippoboscid of the genus *Craterina* (Nováková and Hypša, 2007). Its phylogenetic position indicates that although tsetse flies and hippoboscids are closely related families, their *Sodalis* symbionts have been acquired independently. This arrangement is a typical illustration of sampling significance: in the absence of the *Sitophilus*-symbiont record, the symbionts of glossinids and hippoboscids would form a monophyletic lineage. This finding in turn would imply a common origin of *Sodalis* symbiosis in a pupiparan ancestor. Finally, another *Sodalis* lineage was described by Fukatsu et al. (2007) from the chewing louse, *Columbicola columbae*. The authors pointed out that similar to *Sitophilus*-associated symbionts, these bacteria are likely to display "P-symbiotic traits": they inhabit specialized host cells, their vertical transmission includes migration to the ovary during the host ontogeny, and they are distributed worldwide together with their host species.

Even more dramatic is the history of the genus *Arsenophonus* (Figure 1.2A). This bacterium was described for the first time as a transovarially transmitted infection associated with the son-killer phenomenon in a parasitoid wasp *Nasonia vitripennis* (Huger et al., 1985; Skinner, 1985; Werren et al., 1986). Later, it was formally described as a new genus within the family Enterobacteriacea, closely related to *Proteus* and containing a single species, *Arsenophonus nasoniae* (Gherna et al., 1991). Independent of these studies, the presence of intracellular bacteria infecting various tissues of triatomine bugs was described from microscopic studies (Louis et al., 1986; Hypša, 1993). However, only ten years later, this bacterium was phylogenetically characterized as closely related to *A. nasoniae*, and described as *A. triatominarum* (Hypša and Dale, 1997). In the absence of any other record on *Arsenophonus*, the relatedness of the two bacteria, from *Nasonia* and *Triatoma*, led the authors to discuss possible routes of *Arsenophonus* transmission between triatomines and parasitoid

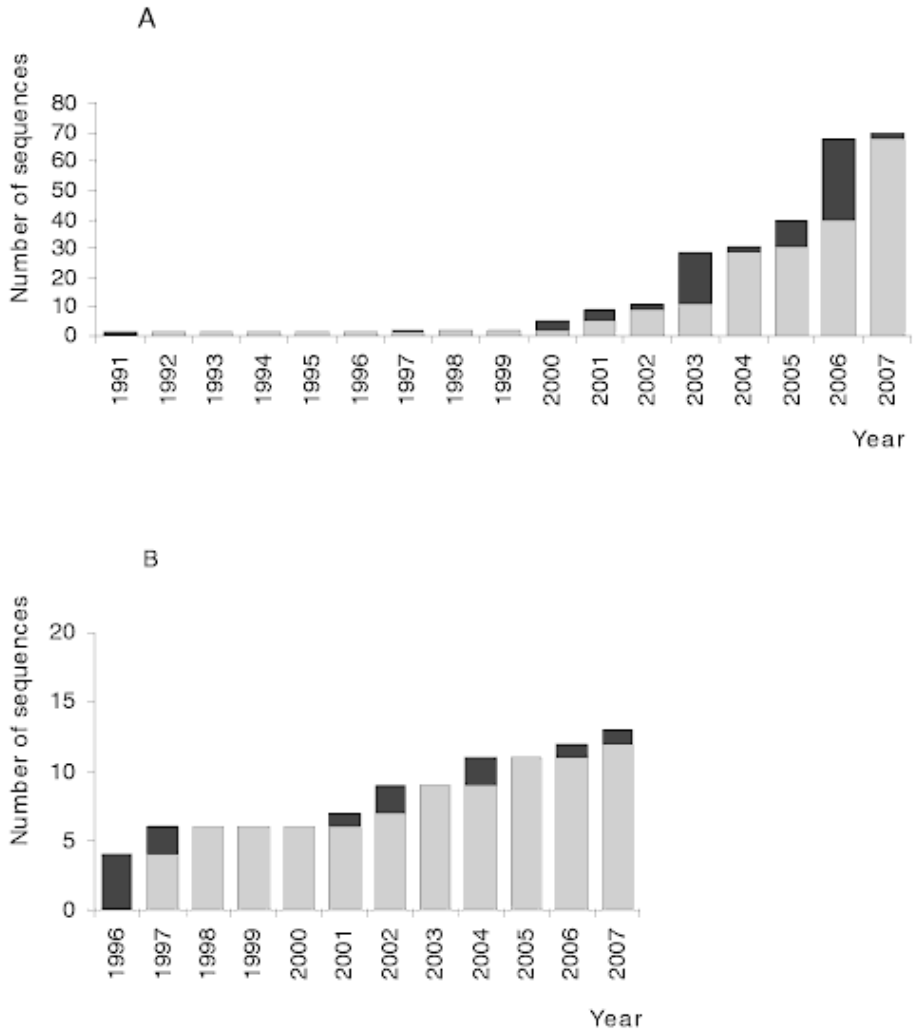


Figure 1.2 An increase of *Arsenophonus* (A) and *Sodalis* (B) records from various insect groups. The bars show cumulative numbers of sequences deposited into the Genbank; dark tops represent new records added in the given year. The sequences are deposited under the following accession numbers: *Arsenophonus*: 1991—M90801; 1997—U91786; 2000—AF263561, AF263562, AF286129; 2001—AF400474, AF400480, AF400478, AY057392; 2002—AY136168, AY136153; 2003—AY265341-AY265348, Y264663-AY264670, AY264673, AY264677; 2004—AY587141, AY587142; 2005—DQ068928, DQ314770-DQ314774, DQ314777, DQ314768, DQ115536; 2006—DQ538372-DQ538379, DQ508171-DQ508186, DQ517447, DQ508193; 2007—EU039464, EU043378; *Sodalis*: 1996—U64867-U648670; 1997—AF005235,Y11391; 2001—AF426460; 2002—AY126638, AY126639; 2004—AY72989, AY729900; 2006—EF174495; 2007—AB303382.

wasps. Since then, the number of *Arsenophonus* records is steadily increasing (Thao et al., 2000a; Trowbridge et al., 2006; Allen et al., 2007) and their current distribution makes any switching scenario meaningless.

The new findings indicate that both genera, *Sodalis* and *Arsenophonus*, may be far more abundant than what can be concluded from their currently known diversity. Moreover, they have given rise to several P-symbiotic lineages that have long been overlooked. Because

symbiont detection and identification almost invariably relies on PCR-based methods, failure to detect a particular bacterium may be a technical consequence of the incapability of the used primers on specimens from these organisms. For example, the frequently used eubacterial primers (O'Neill et al., 1992) were able to detect P-symbionts in several sucking lice, but provided repeatedly negative results with closely related species. The ubiquity of these symbionts in all of the studied species could only be proved by designing specific degenerate primers (Hypša and Křížek, 2007). The danger of mutational changes within priming sites is a notorious problem in molecular biology. It may, however, be commonplace in symbiotic bacteria, where high evolutionary rates and relaxed selection may lead to changes within otherwise conservative regions.

Conclusion

The two last decades of investigations into insect–symbiont associations can be characterized to a great extent by a shift from descriptive research to phylogenetic approach, which instigated a rapid accumulation of molecular data. Due to the unique traits of symbiotic genomes, this research often requires the development and usage of highly specific methods of evolutionary inference. They are represented by various nonstandard evolutionary models (Galtier and Gouy, 1998; Boussau and Gouy, 2006; Gowri-Shankar and Rattray, 2007), complex techniques of coevolutionary analysis (Downie and Gullan, 2005; Takiya et al., 2006), or application of various forms of molecular clock conception (Darby et al., 2001; Hypša and Křížek, 2007). It seems almost ironic that the intensive work into some areas of this research contributes more to the development of phylogenetic methodology than to the solution of actual problems. For example, after many detailed analyses, it is still not clear whether the highly reduced genomes of P-symbionts carry an unequivocal phylogenetic message or whether it has been irreversibly erased by a degradation process. In this respect, it is encouraging to note that a hope has been recently expressed by Wu et al. (2006) that newly sequenced genomes may provide "missing links" that transverse the phylogenetic gaps. They demonstrated this view by finding that when the genome of *Baumannia* is added into the P-symbiotic tree, it forms a branch shorter than usual in other P-symbionts. The focus of several recent studies indicates that the insect–symbiont field may be experiencing another methodological jump. Within the last few years, the increasing number of complete genomes available in the Genbank, and the improved efficiency of computation processes, allowed for various kinds of genome-wide comparisons (Lerat et al., 2003; Comas et al., 2007) and the search for alternative phylogenetic markers (Belda et al., 2005). At the same time, the combination of molecular methods with classical microscopy techniques made it possible to readily connect the phylogenetic data to the histological picture (Moran et al., 2005a; Moran and Dunbar, 2006; Perotti et al., 2006; Fukatsu et al., 2007). This link may be particularly important in connection to the growing number of described symbiotic lineages and the rapidly increasing complexity of symbiotic systems.

Acknowledgments

This work was supported by grant IAA601410708 (Grant Agency of the Academy of Sciences, CZ) and grants LC06073 and MSM 60076605801 (Ministry of Education, CZ).

References

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

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

Aksoy, S., Pourhosseeini, A.A., and Chow, A. (1995). Mycetome endosymbionts of tsetse flies constitute a distinct lineage related to Enterobacteriaceae. *Insect Mol. Biol*. **4:** 5–22.

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

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

Allen, J.M., Reed, D.L., Perotti, M.A., and Braig, H.R. (2007). Evolutionary relationships of "*Candidatus Riesia* spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Appl. Environ. Microbiol*. **73:** 1659–1664.

Anton, A.I., Martinez-Murcia, A.J., and Rodriguez-Valera, F. (1998). Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *J. Mol. Evol*. **47:** 62–72.

Anton, A.I., Martinez-Murcia, A.J., and Rodriguez-Valera, F. (1999). Intraspecific diversity of the 23S rRNA gene and the spacer region downstream in *Escherichia coli*. *J. Bacteriol*. **181:** 2703–2709.

Aschner, M. (1934). Studies on the symbiosis of body louse. Elimination of the symbionts by centrifugation of the eggs. *Parasitology*. **26:** 309–314.

Avise, J.C. (1994). *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Baker, R.H., and DeSalle, R. (1997). Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *System. Biol*. **46:** 654–673.

Baldo, L., Bordenstein, S., Wernegreen, J.J., and Werren, J.H. (2006). Widespread recombination throughout *Wolbachia* genomes. *Mol. Biol. Evol*. **23:** 437–449.

Baldo, L., Prendini, L., Corthals, A., and Werren, J.H. (2007). *Wolbachia* are present in Southern African scorpions and cluster with supergroup F. *Current Microbiol*. **55:** 367–373.

Bandi, C., Sironi, M., Damiani, G., Magrassi, L., Nalepa, C.A., Laudani, U., and Sacchi, L. (1995). The establishment of intracellular symbiosis in an ancestor of cockroaches and termites. *Proc. R. Soc. Lond. Ser. B–Biol. Sci*. **259:** 293–299.

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

Belda, E., Moya, A., and Silva, F.J. (2005). Genome rearrangement distances and gene order phylogeny in gamma-proteobacteria. *Mol. Biol. Evol*. **22:** 1456–1467.

Billoud, B., Guerrucci, M.A., Masselot, M., and Deutsch, J.S. (2000). Cirripede phylogeny using a novel approach: molecular morphometrics. *Mol. Biol. Evol*. **17:** 1435–1445.

Blanquart, S., and Lartillot, N. (2006). A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol. Biol. Evol*. **23:** 2058–2071.

Bordenstein, S., and Rosengaus, R.B. (2005). Discovery of a novel *Wolbachia* supergroup in isoptera. *Current Microbiol*. **51:** 393–398.

Boucher, Y., Douady, C.J., Sharma, A.K., Kamekura, M., and Doolittle, W.F. (2004). Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol*. **186:** 3980–3990.

Boussau, B., and Gouy, M. (2006). Efficient likelihood computations with nonreversible models of evolution. *System. Biol*. **55:** 756–768.

Buchner, P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. John Wiley, New York.

Campbell, B.C., and Purcell, A.H. (1993). Phylogenetic affiliation of BEV, a bacterial parasite of the leafhopper *Euscelidius variegatus*, on the basis of 16S rDNA sequence. *Current Microbiol*. **26:** 37–41.

Canbäck, B., Tamas, I., and Andersson, S.G.E. (2004). A phylogenomic study of endosymbiotic bacteria. *Mol. Biol. Evol*. **21:** 1110–1122.

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J., Werren, J.H., and Bandi, C. (2005). Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. *Microbiol.-Sgm.* **151:** 4015–4022.

Castresana, J., Feldmaier-Fuchs, G., Yokobori, S., Satoh, N., and Paabo, S. (1998). The mitochondrial genome of the hemichordate *Balanoglossus carnosus* and the evolution of deuterostome mitochondria. *Genetics* **150:** 1115–1123.

Charles, H., Heddi, A., and Rahbe, Y. (2001). A putative insect intracellular endosymbiont stem clade, within the Enterobacteriaceae, inferred from phylogenetic analysis based on a heterogeneous model of DNA evolution. *Comptes Rendus Acad. Sci. Ser. III–Sci. Vie–Life Sci.* **324:** 489–494.

Charleston, M.A., and Page, R.D.M. (2002). TreeMap v2.02b; http://taxonomy.zoology.gla.ac.uk/rod/treemap.html.

Charleston, M.A., and Robertson, D.L. (2002). Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *System. Biol.* **51:** 528–535.

Chen, X.A., Li, S., and Aksoy, S. (1999). Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J. Mol. Evol.* **48:** 49–58.

Cilia, V., Lafay, B., and Christen, R. (1996). Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol. Biol. Evol.* **13:** 451–461.

Clark, M.A., Moran, N.A., Baumann, P., and Wernegreen, J.J. (2000). Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* **54:** 517–525.

Clark, M.A., Baumann, L., Thao, M.L.L., Moran, N.A., and Baumann, P. (2001). Degenerative minimalism in the genome of a psyllid endosymbiont. *J. Bacteriol.* **183:** 1853–1861.

Clayton, D.H., Bush, S.E., and Johnson, K.P. (2004). Ecology of congruence: past meets present. *System. Biol.* **53:** 165–173.

Coenye, T., and Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol. Lett.* **228:** 45–49.

Comas, I., Moya, A., Azad, R.K., Lawrence, J.G., and Gonzalez-Candelas, F. (2006). The evolutionary origin of xanthomonadales genomes and the nature of the horizontal gene transfer process. *Mol. Biol. Evol.* **23:** 2049–2057.

Comas, I., Moya, A., and Gonzalez-Candelas, F. (2007). From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic gamma-proteobacteria as a test case. *System. Biol.* **56:** 1–16.

Condon, C., Squirez, C., and Squirez, C.L. (1995). Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* **59:** 623–645.

Dagan, T., and Martin, W. (2007). Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Nat. Acad. Sci.* **104:** 870–875.

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

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

Dale, C., Plague, G.R., Wang, B., Ochman, H., and Moran, N.A. (2002). Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proc. Nat. Acad. Sci. U.S.A.* **99:** 12397–12402.

Dale, C., Beeton, M., Harbison, C., Jones, T., and Pontes, M. (2006). Isolation, pure culture, and characterization of "*Candidatus Arsenophonus arthropodicus*," an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. *Appl. Environ. Microbiol.* **72:** 2997–3004.

Darby, A.C., Birkle, L.M., Turner, S.L., and Douglas, A.E. (2001). An aphid-borne bacterium allied to the secondary symbionts of whitefly. *FEMS Microbiol. Ecol.* **36:** 43–50.

De Vienne, D.M., Giraud, T., and Shykoff, J.A. (2007). When can host shifts produce congruent host and parasite phylogenies? A simulation approach. *J. Evol. Biol.* **20:** 1428–1438.

Degnan, P.H., Lazarus, A.B., Brock, C.D., and Wernegreen, J.J. (2004). Host–symbiont stability and fast evolutionary rates in an ant-bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus Blochmannia*. *System. Biol.* **53:** 95–110.

Dolphin, K., Belshaw, R., Orme, C.D.L., and Quicke, D.L.J. (2000). Noise and incongruence: interpreting results of the incongruence length difference test. *Mol. Phylogenet. Evol.* **17:** 401–406.

Doolittle, W.F., and Bapteste, E. (2007). Inaugural article: pattern pluralism and the tree of life hypothesis. *Proc. Nat. Acad. Sci.* **104:** 2043–2049.

Douglas, A.E. (1989). Mycetocyte symbiosis in insects. *Biol. Rev. Cambridge Philosoph. Soc.* **64:** 409–434.

Downie, D.A., and Gullan, P.J. (2005). Phylogenetic congruence of mealybugs and their primary endosymbionts. *J. Evol. Biol.* **18:** 315–324.

Duron, O., and Gavotte, L. (2007). Absence of *Wolbachia* in nonfilariid worms parasitizing arthropods. *Current Microbiol.* **55:** 193–197.

Eberle, M.W., and McLean, D.L. (1983). Observation of symbiote migration in human body lice with scanning and transmission electron microscopy. *Can J Microbiol* **28:** 755–762.

Farris, J.S., Kallersjo, M., Kluge, A.G., and Bult, C. (1995a). Constructing a significance test for incongruence. *System. Biol.* **44:** 570–572.

Farris, J.S., Kallersjo, M., Kluge, A.G., and Bult, C. (1995b). Testing significance of incongruence. *Cladistics—Int. J. Willi Hennig Soc.* **10:** 315–319.

Favia, G., Ricci, I., Damiani, C., Raddadi, N., Crotti, E., Marzorati, M., Rizzi, A., Urso, R., Brusetti, L., Borin, S., Mora, D., Scuppa, P., Pasqualini, L., Clementi, E., Genchi, M., Corona, S., Negri, I., Grandi, G., Alma, A., Kramer, L., Esposito, F., Bandi, C., Sacchi, L., and Daffonchio, D. (2007). Bacteria of the genus *Asaia* stably associate with *Anopheles stephensi*, an Asian malarial mosquito vector. *Proc. Nat. Acad. Sci. U.S.A.* **104:** 9047–9051.

Foster, P.G. 2004. Modeling compositional heterogeneity. *System. Biol.* **53:** 485–495.

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

Galtier, N., and Gouy, M. (1995). Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Nat. Acad. Sci. U.S.A.* **92:** 11317–11321.

Galtier, N., and Gouy, M. (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* **15:** 871–879.

Garcia-Martinez, J., Martinez-Murcia, A., Anton, A.I., and Rodriguez-Valera, F. (1996). Comparison of the small 16S to 23S intergenic spacer region (ISR) of the rRNA operons of some *Escherichia coli* strains of the ECOR collection and *E. coli* K-12. *J. Bacteriol.* **178:** 6374–6377.

Gaunt, M.W., and Miles, M.A. (2002). An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* **19:** 748–761.

Ge, F., Wang, L.S., and Kim, J. (2005). The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. *PLoS Biol.* **3:** 1709–1718.

Geiger, A., Cuny, G., and 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., Ravel, S., Mateille, T., Janelle, J., Patrel, D., Cuny, G., and Frutos, R. (2007). 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.

Gherna, R.L., Werren, J.H., Weisburg, W., Cote, R., Woese, C.R., Mandelco, L., and Brenner, D.J. (1991). *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *Int. J. Syst. Bacteriol.* **41:** 563–565.

Gil, R., Silva, F.J., Zientz, E., Delmotte, F., Gonzalez-Candelas, F., Latorre, A., Rausell, C., Kamerbeek, J., Gadau, J., Holldobler, B., van Ham, R., Gross, R., and Moya, A. (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Nat. Acad. Sci. U.S.A.* **100:** 9388–9393.

Goldman, N., Anderson, J.P., and Rodrigo, A.G. (2000a). Likelihood-based tests of topologies in phylogenetics. *System. Biol.* **49:** 652–670.

Goldman, N., Anderson, J.P., and Rodrigo, A.G. (2000b). Likelihood-based tests of topologies in phylogenetics. *System. Biol.* **49:** 652–670.

Gowri-Shankar, V., and Rattray, M. (2007). A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. *Mol. Biol. Evol.* **24:** 1286–1299.

Griffiths, G.W., and Beck, S.D. (1974). Effects of antibiotics on intracellular symbionts in pea aphids, *Acyrtosiphon pisum*. *Cell Tissue Res.* **148:** 287–300.

Gruwell, M.E., Morse, G.E., and Normark, B.B. (2007). Phylogenetic congruence of armored scale insects (Hemiptera : Diaspididae) and their primary endosymbionts from the phylum Bacteroidetes. *Mol. Phylogenet. Evol.* **44:** 267–280.

Hasegawa, M., Kishino, H., and Yano, T.A. (1985). Dating of the human–ape splitting by molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22:** 160–174.

Haywood-Farmer, E., and Otto, S.P. (2003). The evolution of genomic base composition in bacteria. *Evolution* **57:** 1783–1792.

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

Herbeck, J.T., Degnan, P.H., and Wernegreen, J.J. (2005). Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (alpha-proteobacteria). *Mol. Biol. Evol.* **22:** 520–532.

Hill, P.D.S., and Campbell, J.A. (1973). The production of symbiont-free *Glossina morsitans* and an associated loss of female fertility. *Trans. R. Soc. Trop. Med. Hyg.* **67:** 727–728.

Hipp, A.L., Hall, J.C., and Sytsma, K.J. (2004). Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. *System. Biol.* **53:** 81–89.

Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M., and Fukatsu, T. (2006). Strict host–symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol.* **4:** 1841–1851.

Huelsenbeck, J.P., Rannala, B., and Larget, B. (2000). A Bayesian framework for the analysis of cospeciation. *Evolution* **54:** 352–364.

Huger, A.M., Skinner, S.W., and Werren, J.H. (1985). Bacterial infections associated with the son-killer trait in the parasitoid wasp *Nasonia* (=Mormoniella) *vitripennis* (Hymenoptera: Pteromalidae). *J. Invertebr. Pathol.* **46:** 272–280.

Hughes, J., Kennedy, M., Johnson, K.P., Palma, R.L., and Page, R.D.M. (2007). Multiple cophylogenetic analyses reveal frequent cospeciation between pelecaniform birds and Pectinopygus lice. *System. Biol.* **56:** 232–251.

Hurst, G.D.D., Hammarton, T.C., Bandi, C., Majerus, T.M.O., Bertrand, D., and Majerus, M.E.N. (1997). The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria. *Genet. Res.* **70:** 1–6.

Hurst, G.D.D., Bandi, C., Sacchi, L., Cochrane, A.G., Bertrand, D., Karaca, I., and Majerus, M.E.N. (1999). *Adonia variegata* (Coleoptera : Coccinellidae) bears maternally inherited Flavobacteria that kill males only. *Parasitology* **118:** 125–134.

Hypša, V. (1993). Endocytobionts of *Triatoma infestans*: distribution and transmission. *J. Invert. Pathol.* **61:** 32–38.

Hypša, V., and Aksoy, S. (1997). Phylogenetic characterization of two transovarially transmitted endosymbionts of the bedbug *Cimex lectularius* (Heteroptera: Cimicidae). *Insect Mol. Biol.* **6:** 301–304.

Hypša, V., and Dale, C. (1997). *In vitro* culture and phylogenetic analysis of "*Candidatus Arsenophonus triatominarum*," an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int. J. Syst. Bacteriol.* **47:** 1140–1144.

Hypša, V., and Křížek, J. (2007). Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura). *Microb. Ecol.* **54:** 242–251

Ishikawa, H. (1982). Isolation of the intracellular symbionts and partial characterization of their RNA species of the elder aphid *Acyrtosiphon magnoliae*. *Comp Biochem. Physiol.* **72:** 239–247.

Itoh, T., Martin, W., and Nei, M. (2002). Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Nat. Acad. Sci. U.S.A.* **99:** 12944–12948.

Johnson, K.P., Drown, D.M., and Clayton, D.H. (2001). A data based parsimony method of cophylogenetic analysis. *Zoologica Scripta* **30:** 79–87.

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequences data, and the branching order in hominoidea. *J. Mol. Evol.* **29:** 170–179.

Koga, R., Tsuchida, T., and Fukatsu, T. (2003). Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont Buchnera in an aphid. *Proc. R. Soc. Lond. Ser. B–Biol. Sci.* **270:** 2543–2550.

Korbel, J.O., Snel, B., Huynen, M.A., and Bork, P. (2002). SHOT: a web server for the construction of genome phylogenies. *Trends Genet.* **18:** 158–162.

Lake, J.A. (1994). Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Nat. Acad. Sci. U.S.A.* **91:** 1455–1459.

Lambert, J.D., and Moran, N.A. (1998). Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc. Nat. Acad. Sci. U.S.A.* **95:** 4458–4462.

Larget, B., Simon, D.L., Kadane, J.B., and Sweet, D. (2005). A Bayesian analysis of metazoan mitochondrial genome arrangements. *Mol. Biol. Evol.* **22:** 486–495.

Lavrov, D.V., Brown, W.M., and Boore, J.L. (2004). Phylogenetic position of the Pentastomida and (pan)crustacean relationships. *Proc. R. Soc. Lond. Ser. B–Biol. Sci.* **271:** 537–544.

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

Lerat, E., Daubin, V., and Moran, N.A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-proteobacteria. *PLoS Biol.* **1:** 101–109.

Lerat, E., Daubin, V., Ochman, H., and Moran, N.A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3:** 807–814.

Lin, C.K., Hung, C.L., Chiang, Y.C., Lin, C.M., and Tsen, H.Y. (2004). The sequence heterogenicities among 16S rRNA genes of *Salmonella* serovars and the effects on the specificity of the primers designed. *Int. J. Food Microbiol.* **96:** 205–214.

Lo, N., Bandi, C., Watanabe, H., Nalepa, C., and Beninati, T. (2003). Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol. Biol. Evol.* **20:** 907–913.

Lockhart, P.J., Steel, M.A., Hendy, M.D., and Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence. *Mol. Biol. Evol.* **11:** 605–612.

Lopez-Vaamonde, C., Rasplus, J.Y., Weiblen, G.D., and Cook, J.M. (2001). Molecular phylogenies of fig wasps: partial cocladogenesis of pollinators and parasites. *Mol. Phylogenet. Evol.* **21:** 55–71.

Louis, C., Drif, L., and Vago, C. (1986). Mise en évidence et étude ultrastructurale de procaryotes de type rickettsien dans les glandes salivaires des Triatomidae (Heteroptera) = Evidence and ultrastructural study of *Rickettsia*-like prokaryotes in salivary glands of Triatomidae (Heteroptera) *Ann. Soc. Entomol. Fr.* **22:** 153–162.

Luz, S.P., Rodriguez-Valera, F., Lan, R.T., and Reeves, P.R. (1998). Variation of the ribosomal operon 16S-23S gene spacer region in representatives of *Salmonella enterica* subspecies. *J. Bacteriol.* **180:** 2144–2151.

Marchandin, H., Teyssier, C., de Buochberg, M.S., Jean-Pierre, H., Carriere, C., and Jumas-Bilak, E. (2003). Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: implications for phylogeny and taxonomy. *Microbiol.-Sgm.* **149:** 1493–1501.

Martinez-Murcia, A.J., Anton, A.I., and Rodriguez-Valera, F. (1999). Patterns of sequence variation in two regions of the 16S rRNA multigene family of *Escherichia coli*. *Int. J. Syst. Bacteriol.* **49:** 601–610.

Marzorati, M., Alma, A., Sacchi, L., Pajoro, M., Palermo, S., Brusetti, L., Raddadi, N., Balloi, A., Tedeschi, R., Clementi, E., Corona, S., Quaglino, F., Bianco, P.A., Beninati, T., Bandi, C., and Daffonchio, D. (2006). A novel bacteroidetes symbiont is localized in *Scaphoideus titanus*, the insect vector of flavescence doree in *Vitis vinifera*. *Appl. Environ. Microbiol.* **72:** 1467–1475.

Merkle, D., and Middendorf, M. (2005). Reconstruction of the cophylogenetic history of related phylogenetic trees with divergence timing information. *Theory Biosci.* **123:** 277–299.

Mooers, A.O., and Holmes, E.C. (2000). The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* **15:** 365–369.

Moran, N.A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Nat. Acad. Sci. U.S.A.* **93:** 2873–2878.

Moran, N.A., Dale, C., Dunbar, H., Smith, W.A., and Ochman, H. (2003). Intracellular symbionts of sharpshooters (Insecta : Hemiptera : Cicadellinae) form a distinct clade with a small genome. *Environ. Microbiol.* **5:** 116–126.

Moran, N.A., and Dunbar, H.E. (2006). Sexual acquisition of beneficial symbionts in aphids. *Proc. Nat. Acad. Sci. U.S.A.* **103:** 12803–12806.

Moran, N.A., Munson, M.A., Baumann, P., and Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. Ser. B–Biol. Sci.* **253:** 167–171.

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

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

Moran, N.A., and Wernegreen, J.J. (2000). Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol. Evol.* **15:** 321–326.

Munson, M.A., Baumann, P., Clark, M.A., Baumann, L., Moran, N.A., Voegtlin, D.J., and Campbell, B.C. (1991a). Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *J. Bacteriol.* **173:** 6321–6324.

Munson, M.A., Baumann, P., and Kinsey, M.G. (1991b). *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *Int. J. Syst. Bacteriol.* **41:** 566–568.

Munson, M.A., Baumann, P., and Moran, N.A. (1992). Phylogenetic relationships of the endosymbionts of mealybugs (Homoptera: Pseudococcidae) based on 16S rDNA sequences. *Mol. Phylogenet. Evol.* **1:** 26–30.

Murrell, A., Campbell, N.J.H., and Barker, S.C. (2003). The value of idiosyncratic markers and changes to conserved tRNA sequences from the mitochondrial genome of hard ticks (Acari : Ixodida : Ixodidae) for phylogenetic inference. *System. Biol.* **52:** 296–310.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. (2006). The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314:** 267–267.

Negrisolo, E., Minelli, A., and Valle, G. (2004). Extensive gene order rearrangement in the mitochondrial genome of the centipede *Scutigera coleoptrata*. *J. Mol. Evol.* **58:** 413–423.

Nogge, G. (1978). Aposymbiotic tsetse flies, *Glossna morsitans morsitans*, obtained by feeding on rabbits immunized specifically with symbionts. *J. Insect Physiol.* **24:** 299–304.

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

Ochman, H., and Wilson, A.C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes *J. Mol. Evol.* **26:** 74–86.

O'Neill, S.L., Giordano, R., Colbert, A.M.E., Karr, T.L., and Robertson, H.M. (1992). 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proc. Nat. Acad. Sci. U.S.A.* **89:** 2699–2702.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276:** 734–740.

Page, R.D.M. (ed.). (2003). *Tangled Trees: Phylogeny, Cospeciation, and Coevolution*. University of Chicago Press, Chicago.

Paraskevopoulos, C., Bordenstein, S.R., Wernegreen, J.J., Werren, J.H., and Bourtzis, K. (2006). Toward a *Wolbachia* multilocus sequence typing system: discrimination of *Wolbachia* strains present in *Drosophila* species. *Curr. Microbiol.* **53:** 388–395.

Percy, D.M., Page, R.D.M., and Cronk, Q.C.B. (2004). Plant-insect interactions: double-dating associated insect and plant lineages reveals asynchronous radiations. *System. Biol.* **53:** 120–127.

Perotti, M.A., Clarke, H.K., Turner, B.D., and Braig, H.R. (2006). *Rickettsia* as obligate and mycetomic bacteria. *Faseb J.* **20:** 2372–2374.

Pettersson, B., Bolske, G., Thiaucourt, F., Uhlen, M., and Johansson, K.E. (1998). Molecular evolution of *Mycoplasma capricolum* subsp. capripneumoniae strains, based on polymorphisms in the 16S rRNA genes. *J. Bacteriol.* **180:** 2350–2358.

Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Muller, M., and Le Guyader, H. (2000). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. Lond. Ser. B–Biol. Sci.* **267:** 1213–1221.

Podsiadlowski, L., Kohlhagen, H., and Koch, M. (2007). The complete mitochondrial genome of *Scutigerella causeyae* (Myriapoda : Symphyla) and the phylogenetic position of *Symphyla. Mol. Phylogenet. Evol.* **45:** 251–260.

Posada, D., and Buckley, T. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *System. Biol.* **53:** 793–808.

Puchta, O. (1955). Experimentelle untersuchungen über die bedeutung der symbiose der kleiderlaus *Pediculus vestimenti* burm. *Z. Parastenkd.* **17:** 1–40.

Quicke, D.L.J., Jones, O.R., and Epstein, D.R. (2007). Correcting the problem of false incongruence due to noise imbalance in the incongruence length difference (ILD) test. *System. Biol.* **56:** 496–503.

Ramirez, M.J. (2006). Further problems with the incongruence length difference test: "hypercongruence" effect and multiple comparisons. *Cladistics* **22:** 289–295.

Reed, D.L., and Hafner, M.S. (2002). Phylogenetic analysis of bacterial communities associated with ectoparasitic chewing lice of pocket gophers: a culture-independent approach. *Microb. Ecol.* **44:** 78–93.

Reeves, W.K., Loftis, A.D., Gore, J.A., and Dasch, G.A. (2005). Molecular evidence for novel *Bartonella* species in *Trichobius major* (Diptera : Streblidae) and *Cimex adjunctus* (Hemiptera : Cimicidae) from two southeastern bat caves, USA. *J. Vector Ecol.* **30:** 339–341.

Rokas, A., and Holland, P.W.H. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15:** 454–459.

Ronquist, F. (1995). Reconstructing the history of host–parasite associations using generalised parsimony. *Cladistics—Int. J. Willi Hennig Soc.* **11:** 73–89.

Sakurai, M., Koga, R., Tsuchida, T., Meng, X.Y., and Fukatsu, T. (2005). Rickettsia symbiont in the pea aphid *Acyrthosiphon pisum*: novel cellular tropism, effect on host fitness, and interaction with the essential symbiont *Buchnera*. *Appl. Environ. Microbiol.* **71:** 4069–4075.

Sauer, C., Stackebrandt, E., Gadau, J., Holldobler, B., and Gross, R. (2000). Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus Blochmannia* gen. nov. *Int. J. Sys.. Evol. Microbiol.* **50:** 1877–1886.

Schröder, D., Deppisch, H., Obermayer, M., Krohne, G., Stackebrandt, E., Holldobler, B., Goebel, W., and Gross, R. (1996). Intracellular endosymbiotic: Bacteria of *Camponotus* species (carpenter ants): systematics, evolution and ultrastructural characterization. *Mol. Microbiol.* **21:** 479–489.

Shi, X., Gu, H., Susko, E., and Field, C. (2005). The comparison of the confidence regions in phylogeny. *Mol. Biol. Evol.* **22:** 2285–2296.

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

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16:** 1114–1116.

Sironi, M., Bandi, C., Sacchi, L., DiSacco, B., Damiani, G., and Genchi, C. (1995). Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. *Mol. Biochem. Parasitol.* **74:** 223–227.

Snel, B., Huynen, M.A., and Dutilh, B.E. (2005). Genome trees and the nature of genome evolution. *Ann. Rev. Microbiol.* **59:** 191–209.

Spaulding, A.W., and von Dohlen, C.D. (1998). Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera : Sternorrhyncha). *Mol. Biol. Evol.* **15:** 1506–1513.

Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. *Science* **297:** 89–91.

Steel, M. (2005). Should phylogenetic models be trying to "fit an elephant"? *Trends Genet.* **21:** 307–309.

Subandiyah, S., Nikoh, N., Tsuyumu, S., Somowiyarjo, S., and Fukatsu, T. (2000). Complex endosymbiotic microbiota of the citrus psyllid *Diaphorina citri* (Homoptera : Psylloidea). *Zool. Science* **17:** 983–989.

Susko, E., Leigh, J., Doolittle, W.F., and Bapteste, E. (2006). Visualizing and assessing phylogenetic congruence of core gene sets: a case study of the gamma-proteobacteria. *Mol. Biol. Evol.* **23:** 1119–1030.

Swain, T.D., and Taylor, D.J. (2003). Structural rRNA characters support monophyly of raptorial limbs and paraphyly of limb specialization in water fleas. *Proc. R. Soc. Lond. Ser. B–Biol. Sci.* **270:** 887–896.

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

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* **9:** 678–687.

Thao, M.L., and Baumann, P. (2004a). Evidence for multiple acquisition of Arsenophonus by whitefly species (Sternorrhyncha : Aleyrodidae). *Curr. Microbiol.* **48:** 140–144.

Thao, M.L., and Baumann, P. (2004b). Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl. Environ. Microbiol.* **70:** 3401–3406.

Thao, M.L., Clark, M.A., Baumann, L., Brennan, E.B., Moran, N.A., and Baumann, P. (2000a). Secondary endosymbionts of psyllids have been acquired multiple times. *Curr. Microbiol.* **41:** 300–304.

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

Thao, M.L., Moran, N.A., Abbot, P., Brennan, E.B., Burckhardt, D.H., and Baumann, P. (2000b). Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Appl. Environ. Microbiol.* **66:** 2898–2905.

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

Trowbridge, R.E., Dittmar, K., and Whiting, M.F. (2006). Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult *Hippoboscidae* and *Streblidae* (Hippoboscoidea). *J. Invert. Pathol.* **91:** 64–68.

Tsuchida, T., Koga, R., Sakurai, M., and Fukatsu, T. (2006). Facultative bacterial endosymbionts of three aphid species, *Aphis craccivora*, *Megoura crassicauda* and *Acyrthosiphon pisum*, sympatrically found on the same host plants. *Appl. Entomol. Zool.* **41:** 129–137.

Vaishampayan, P.A., Dhotre, D.P., Gupta, R.P., Lalwani, P., Ghate, H., Patole, M.S., and Shouche, Y.S. (2007). Molecular evidence and phylogenetic affiliations of *Wolbachia* in cockroaches. *Mol. Phylogenet. Evol.* **44:** 1346–1351.

Weinert, L.A., Tinsley, M.C., Temperley, M., and Jiggins, F.M. (2007). Are we underestimating the diversity and incidence of insect bacterial symbionts? A case study in ladybird beetles. *Biol. Lett.* 3: **678–681.**

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

Werren, J.H. (1997). Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42:** 587–609.

Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., Moran, N.A., and Eisen, J.A. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4:** 1079–1092.

Yang, Z.H., and Roberts, D. (1995). On the use of nucleotide acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* **12:** 451–458.

Yap, W.H., Zhang, Z.S., and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* **181:** 5201–5209.

Yen, J.J., and Barr, A.R. (1971). New hypothesis for the case of cytoplasmic incompatibility in *Culex pipiens*. *Nature* **232:** 657–658.

Yoder, A.D., Irwin, J.A., and Payseur, B.A. (2001). Failure of the ILD to determine data combinability for slow loris phylogeny. *System. Biol.* **50:** 408–424.

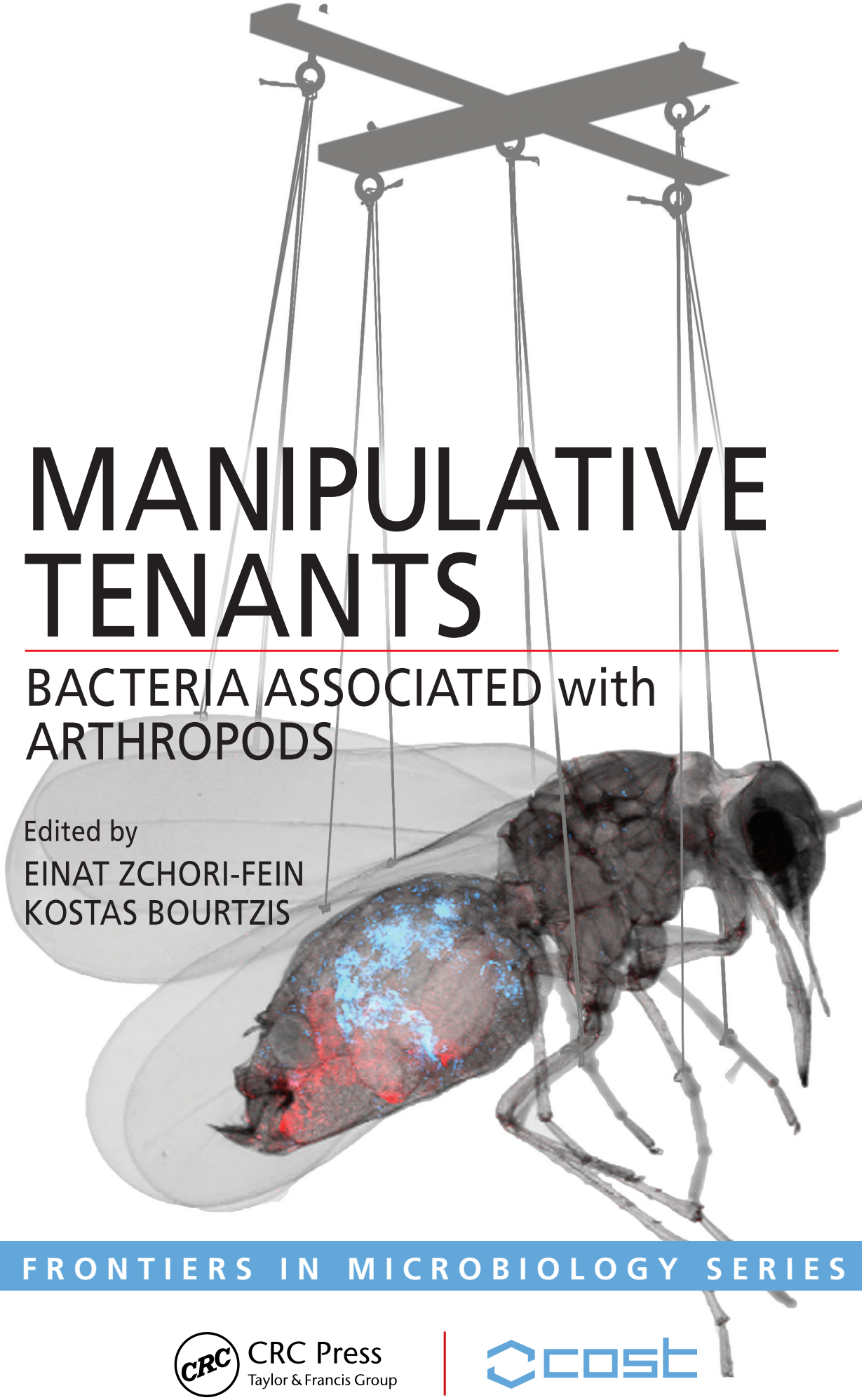
Zchori-Fein, E., and Brown, J.K. (2002). Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae). *Ann. Entomol. Soc. Amer.* **95:** 711–718.

Zchori-Fein, E., and Perlman, S.J. (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Mol. Ecol.* **13:** 2009–2016.

Zhou, W.G., Rousset, F., and O'Neill, S. (1998). Phylogeny and PCR-based classification of *Wolbachia* strains using wsp gene sequences. *Proc. R. Soc. Lond. Ser. B–Biol. Sci.* **265:** 509–515.

Book Chapter 2:

Wilkes, T. E., Duron, O., Darby, A. C., Hypša, V., **Nováková, E.,** & Hurst, G. D. (2011). 12 The Genus *Arsenophonus*. *Manipulative tenants: bacteria associated with arthropods*, 225.



MANIPULATIVE TENANTS

BACTERIA ASSOCIATED with ARTHROPODS

Edited by
EINAT ZCHORI-FEIN
KOSTAS BOURTZIS

FRONTIERS IN MICROBIOLOGY SERIES

CRC Press
Taylor & Francis Group

cost

MANIPULATIVE TENANTS

BACTERIA ASSOCIATED with ARTHROPODS

Edited by
EINAT ZCHORI-FEIN
KOSTAS BOURTZIS

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Cover design: Mirit Assaf

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20110705

International Standard Book Number-13: 978-1-4398-2750-5 (eBook - PDF)

12 The Genus *Arsenophonus*

Timothy E. Wilkes, Olivier Duron, Alistair C. Darby, Václav Hypša, Eva Nováková, and Gregory D. D. Hurst

CONTENTS

Introduction to the Genus *Arsenophonus* 226
Biodiversity within the Genus *Arsenophonus* 227
Hosts Infected and Nature of the Symbioses Found 230
Arsenophonus nasoniae/Pteromalid Wasps 231
Microbial Facts 231
Interaction with Host 231
Arsenophonus arthropodicus/Hippoboscid Flies 234
Microbial Facts 234
Interaction with Host 234
Arsenophonus triatominarum/Triatomine Bugs 235
Microbial Facts 235
Interaction with Host 235
Riesia pediculicola/Lice 237
Microbial Facts 237
Interaction with Host 237
Phlomobacter fragariae/Planthoppers/Plants 238
Microbial Features 238
Interaction with Host 238
Arsenophonus Genomes 239
The Genome of *A. nasoniae* 239
The Genome of *R. pediculicola* 240
Prospects 240
References 241

Genus	*Arsenophonus*
Type species	*nasoniae*
Family	Gamma-Proteobacteria
Order	Enterobacteriacae
Description year	1991
Origin of name	*Arsen* = male, *phonus* = slayer
Description	
Reference	Gherna et al. (1991)

INTRODUCTION TO THE GENUS ARSENOPHONUS

In 1985, individual lines of the wasp *Nasonia vitripennis* were observed to produce strongly female-biased broods, containing just 0–5% sons (Skinner 1985). The trait was maternally inherited, and the sex ratio bias found to be associated with the death of male embryos. Pertinent to this book, the trait was additionally found to be infectiously transmitted between wasp lines following superparasitism. This latter observation implied the presence of an infectious agent. Detailed microbiological study led to the isolation of a new bacterium in liquid culture, for which Koch's postulates were then fulfilled (Werren et al. 1986). The bacterium was formally described as *Arsenophonus nasoniae* sp. nov. gen. nov., the type species of the genus *Arsenophonus* (Gherna et al. 1991). The formal status of the genus was established by Werren in 2005 (Werren 2005).

Since its initial discovery, membership of the genus *Arsenophonus* has expanded, and it is now known to be both widespread and biologically diverse. In terms of incidence it is one of the "big four" inherited symbionts of arthropods, being present in ca. 5% of species. Infections have been described in a diverse range of arthropod hosts, including arachnids, ticks, cockroaches, hemipterans, hymenopterans, lice, flies, and coleopterans (Duron et al. 2008). In terms of symbiotic diversity, several different stages of symbiosis can be found within the genus, and interactions include facultative and obligate, parasitic and beneficial. The type species *A. nasoniae* is a parasite with a large genome with substantial metabolic capability (Darby et al. 2010), and can be grown in supplemented cell-free media (Werren et al. 1986). Other strains have reduced genomes, can be grown only in cell culture, and are probably beneficial for host fitness. *Riesia*, the primary symbiont of lice, can be considered a biologically highly derived species of *Arsenophonus*. It has a very diminished genome size (Kirkness et al. 2010), is unculturable, and is required by the host. This makes the *Arsenophonus* clade one where all stages of symbiosis can be identified, and in which we can potentially reconstruct the evolutionary processes that underlie changes in symbiotic relationships.

One other point of note is the placement of the genus compared to other microbes. The genus *Arsenophonus* falls in the gamma subdivision of the Proteobacteria (Figure 12.1). As such, it benefits from possessing a plethora of well-studied "model" organism comparators, including *E. coli*, human pathogens such as *Yersinia pestis*

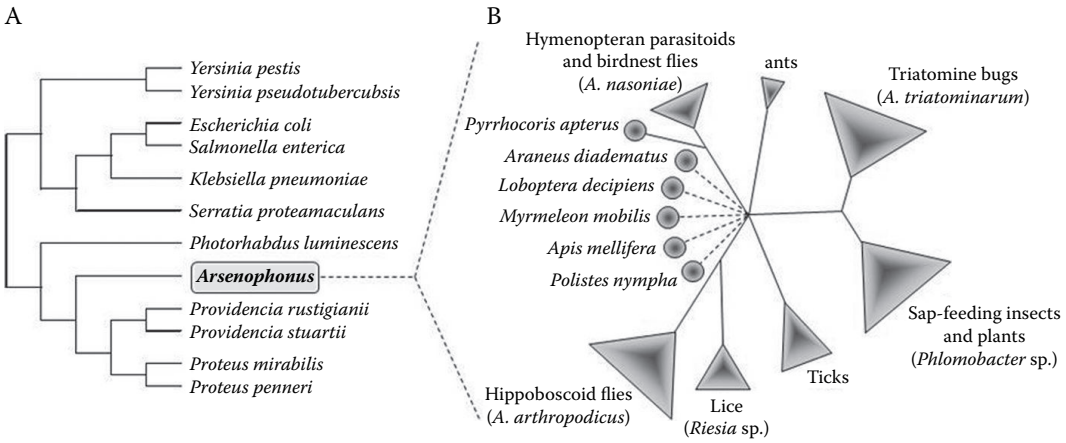


FIGURE 12.1 Phylogeny of *Arsenophonus*. (A) The phylogenetic relationships between *Arsenophonus* and representative members of the Enterobacteriaceae family. (B) An unrooted phylogenetic tree of the main subdivisions of *Arsenophonus*. Dotted lines represent *Arsenophonus* strains for which relationships with the other strains are uncertain. Triangles represent described diversity within each lineage based on a single *Arsenophonus* strain. Circles represent single strains.

and *Salmonella typhimurium*, and insect pathogens such as *Photorhabdus luminescens*. These each provide good hypotheses for gene function and capabilities in the *Arsenophonus* genome, where functional work has previously been lacking. There are also a variety of well-understood insect symbionts to which it can be compared, including facultative symbionts such as *Hamiltonella defensa* and *Sodalis*, and primary symbionts like *Buchnera* and *Wigglesworthia*. Its position within the gamma-Proteobacteria is now well resolved from multigene data, with *Proteus* and *Providencia* clearly the closest allied genera (Darby et al. 2010).

Arsenophonus therefore represents a symbiotic clade of great potential interest. In this chapter, we first describe the known microbial biodiversity within the genus *Arsenophonus* as it is currently recognized, and debates about membership of it. We then summarize the current state of knowledge of the five *Arsenophonus*-host interactions that are best characterized, and review recent findings from draft genome sequences of two members of the clade, *A. nasoniae* and *Riesia pediculicola*. Finally, we point to four areas of future research for workers on this clade of symbionts.

BIODIVERSITY WITHIN THE GENUS *ARSENOPHONUS*

Within the genus *Arsenophonus*, there are currently two species formally described beyond *A. nasoniae*, the type. These are Cand. *Arsenophonus triatominarum*, isolated from triatomine bugs, and Cand. *Arsenophonus arthropodicus*, isolated from the pigeon louse fly, *Pseudolynchia canariensis*. There has been a recent debate about inclusion of other bacteria, notably Cand. *Phlomobacter frageriae* and Cand. *Riesia pediculicola*.

Cand. *Phlomobacter fragariae*, the causative agent of marginal chlorosis in strawberry plants and endosymbiont of the planthopper *Cixius wagneri*, was described by Danet et al. (2003). In 2005, Werren (2005) suggested *Phlomobacter* fell within the clade *Arsenophonus*. Subsequent phylogenetic analysis using the sequence of the 16S rRNA gene has indeed indicated that this bacterium falls within the genus *Arsenophonus*, and being junior to it in description, should be incorporated within it, and it is now regarded as a member of the clade *Arsenophonus* (Bressan et al. 2009). Cand. *Riesia pediculicola*, a genus of symbionts found in lice, is a very different bacterium from described *Arsenophonus* species in terms of symbiosis, microbiology, and genomic content. Notwithstanding the clear biological differentiation of the microbe, *Riesia* falls phylogenetically within the genus *Arsenophonus* rather than as a distinct clade. Without prejudging this debate, we include *Riesia* in the chapter by virtue of the interestingly different nature of this related bacterium.

Beyond these described species, *Arsenophonus* strains are generally known only from the 16S rRNA gene sequence recovered from insect specimens using general eubacterial primers (see Table 12.1). In some cases, such as *Arsenophonus* in *Bemisia*, fluorescence *in situ* hybridization (FISH) reveals the infection to be present within bacteriocytes inside the insect (Gottlieb et al. 2008). In most other cases, presence is indicated solely by a polymerase chain reaction (PCR) positive assay using template material derived from whole arthropods. While there is no direct confirmation of symbiosis in these cases, it is highly likely that these represent symbiotic partners of the insects in question.

The incidence of *Arsenophonus* is best estimated from a "blind" survey. Duron et al. (2008) used a PCR assay for *Arsenophonus* to investigate how commonly it occurred in field-collected insects. They reported infection in 6 of 136 species, representing an estimated incidence of 5%. The survey concluded there was no obvious bias in the taxa infected, and it is notable the same survey did conclude that *Cardinium*, *Wolbachia*, and *Spiroplasma ixodetis* were more common in spiders than insects, a pattern not found for *Arsenophonus*. The survey data provide a slightly different picture from individual records, which suggests *Arsenophonus* infections may be most common among species of Hemiptera.

Study of the biodiversity of *Arsenophonus* strains has traditionally derived solely from the 16S rRNA gene sequence. However, this marker may perform poorly in resolving patterns of relatedness in the genus, and the interpretation of past investigations is now subject to caution since several recent studies have revealed major failures in the framework for studies of *Arsenophonus* phylogenetics (Novakova et al. 2009; Sorfova et al. 2008). Although 16S rRNA gene sequence marker identifies *Arsenophonus* as a robust monophyletic clade, Novakova et al. (2009) demonstrated that inner topology is unstable, mainly because of methodological artifacts due to insufficient sequence acquisition, preventing clear understanding of the evolutionary relationships between *Arsenophonus* strains. In parallel, Sorfova et al. (2008) showed that several *Arsenophonus* strains carry a high number of rRNA gene copies in their genomes, a characteristic also reported in other Enterobacteriaceae genomes (Moran et al. 2008). Variability between individual rRNA gene copies within a bacterium has serious consequences for phylogenetic inferences, as it depicts inexact evolutionary trajectories for *Arsenophonus* infections.

TABLE 12.1
Details of Currently Known Members of the *Arsenophonus* Clade

Host Organism	Strain Nomenclature	Phenotype	Reference
Arachnida			
Ixodida			
Amblyomma americanum (Ixodidae)	*Arsenophonus* sp.	Not known	Clay et al. 2008
Dermacentor variabilis (Ixodidae)	*Arsenophonus* sp.	Not known	Grindle et al. 2003
Dermacentor andersoni	*Arsenophonus* sp.	Not known	Dergousoff and Chilton 2010
Aranea			
Araneus diadematus (Araneidae)	*Arsenophonus* sp.	Not known	Duron et al. 2008
Insecta			
Blattaria			
Loboptera decipiens (Blatellidae)	*Arsenophonus* sp.	Not known	Duron et al. 2008
Diptera			
Protocalliphora azurea (Calliphoridae)	*Arsenophonus* sp.	Not known	Duron et al. 2008
Various hippoboscid species (Hippoboscidae)[a]	*A. arthropodicus*, *Arsenophonus* sp.	Not known	Dale et al. 2006 Duron et al. 2008 Novakova et al. 2009
Various streblid species (Streblidae)[a]	*Arsenophonus* sp.	Not known	Novakova et al. 2009 Trowbridge et al. 2006
Hemiptera			
Pyrrhocoris apterus (Pyrrhocoridae)	*Arsenophonus* sp.	Not known	Duron et al. 2008
Various triatomine species (Reduvidae)[a]	*A. triatominarum*, *Arsenophonus* sp.	Not known	Hypsa and Dale 1997 Sorfova et al. 2008
Various planthopper species (Cixiidae)[a,b]	*Phlomobacter* sp.	Phytopathogen	Bressan et al. 2008 Danet et al. 2003 Semetey et al. 2007
Nilaparvata lugens (Dephacidae)	*Arsenophonus* sp.	Not known	Wang et al. 2010
Various whiteflies species (Aleyrodidae)[a]	*Arsenophonus* sp.	Bacteriocytes associated	Gottlieb et al. 2008 Thao and Baumann 2004 Zchori-Fein and Brown 2002
Various aphid species (Aphididae)[a]	*Arsenophonus* sp.	Not known	Najar-Rodriguez et al. 2009 Russell et al. 2003
Various psyllid species (Psylloidea)[a]	*Arsenophonus* sp.	Not known	Hansen et al. 2007 Spaulding and von Dohlen 2001 Subandiyah et al. 2000 Thao et al., 2000a 2000b

(*continued*)

TABLE 12.1 (continued)
Details of Currently Known Members of the *Arsenophonus* Clade

Host Organism	Strain Nomenclature	Phenotype	Reference
Hymenoptera			
Nasonia vitripennis (Pteromelidae)	*A. nasoniae*	Son killer	Gherna et al. 1991 Werren et al. 1986 Skinner 1985 Huger et al. 1985
Polistes nympha (Vespidae)	*Arsenophonus* sp.	Not known	Duron et al. 2008
Apis mellifera (Apidae)	*Arsenophonus* sp.	Not known	Babendreier et al. 2007
Two ant species (Formicidae)	*Arsenophonus* sp.	Not known	Novakova et al. 2009
Neuroptera			
Myrmeleon mobilis (Myrmeleontidae)	*Arsenophonus* sp.	Not known	Dunn and Stabb 2005
Phthiraptera			
Various lice species (Pediculidea)[a]	*Riesia* sp.	Bacteriocytes associated	Allen et al. 2007 Sasaki-Fukatsu et al. 2006 Perotti et al. 2007
	Plants		
Beta vulgaris (sugar beet)[b]	*Phlomobacter* sp., SBR bacterium	Phytopathogen	Bressan et al. 2008 Semetey et al. 2007
Fragaria sp. (strawberry)[b]	*Phlomobacter fragariae*	Phytopathogen	Danet et al. 2003 Zreik et al. 1998

[a] Most of the host species of these families are infected and are not detailed exhaustively.
[b] The same *Arsenophonus* (*Phlomobacter*) strains are found in Cixidae hosts and plants.

In light of this, more recent studies of biodiversity have utilized a multilocus system for approaching phylogeny in the clade based on three housekeeping genes. These conformed to the basic properties of a multilocus strain typing (MLST) system—single copy, no internal recombination, with the majority of strains (save *Riesia*) amplifiable with conserved primers (Table 12.2). The sequences of these genes have been used to examine patterns of transmission of *Arsenophonus* between species, particularly *A. nasoniae* in the guild of filth fly parasites (Duron et al. 2010).

HOSTS INFECTED AND NATURE OF THE SYMBIOSES FOUND

Information about the nature of these symbioses is known for a few cases that we detail below. Basic comparative information about these symbionts is given in Table 12.3. The details of these symbioses reveal varying importance of horizontal

TABLE 12.2
Oligonucleotide Primer Sequences, PCR Annealing Temperature and Expected Amplicon Size for a Variety of Genetic Markers in *Arsenophonus* Used in Reconstructing the Relatedness of Strains

Gene	Hypothetical Product	Primers (5′–3′)		Tm	Fragment Size
16S rRNA	Small ribosomal subunit	ArsF ArsR2	GGGTTGTAAAGTACTTTCAGTCGT GTAGCCCTRCTCGTAAGGGCC	52°C	804 bp
fbaA	Fructose-bisphosphate aldolase class II	fbaAf fbaAr	GCYGCYAAAGTTCRTTCTCC CCWGAACCDCCRTGGAAAACAAAA	52°C	659 bp
ftsK	Cell division protein (DNA translocase)	ftsKf ftsKr	GTTGTYATGGTYGATGAATTTGC GCTCTTCATCACYATCAWAACC	52°C	445 bp
yaeT	Outer membrane protein assembly factor	yaeTf yaeTr	GCATACGGTTCAGACGGGTTTG GCCGAAACGCCTTCAGAAAAG	52°C	473 bp
zapA	Zinc-dependent metalloprotease	zapAf zapAr	GGGTCACATACCTATTTT GTAGTCGCCTGGGTGGG	50°C	594 bp
aprA	Zinc-dependent metalloprotease	aprAf aprAr	CATTTAATTCCAAGAAC GAAAGTCTGCTTGTCCATCTCC	50°C	513 bp

Note: zapA and aprA products only obtained in *A. nasoniae.*

and vertical transmission of the symbiont, the presence of sex ratio distorting activity in one strain (but not others), potential beneficial effects in some, and confirmed beneficial effects associated with integration into host anatomy in others.

Arsenophonus Nasoniae/Pteromalid Wasps

Microbial Facts

Rod-like bacterium, 6.9–10.0 μm in length and 0.40–0.57 μm in diameter (Gherna et al. 1991). Possesses a genome of c. 3.5 MB with at least one small and one large plasmid (Darby et al. 2010). Culturable on cell-free GC media with addition of Kellogg's supplement, where it produces colonies in 4–5 days (Werren et al. 1986). *In vitro* growth generally conducted at 25–30°C. May be cryopreserved in glycerol stocks.

Interaction with Host

Widely disseminated infection found throughout host tissues (Huger et al. 1985). Infection induces a female-biased sex ratio in its host, associated with mortality of 80% of male offspring (Skinner 1985).

TABLE 12.3
Summary of Information on the Five Most Studied Members of the Genus *Arsenophonus*

	Microbial Morphology (Length, Diameter)	Grows in Cell-Free Culture?	Maintained in Insect Cell Culture?	Genome Size	Genome Structure	Known Phenotypes
A. nasoniae	Rod 7–10 × 0.4–0.6 µm	Y	Y	~3.5 MB	Chromosome + small plasmid + 1 or more large plasmids	Son killing
C. A. triatominarum	Rod 15 × 1.5 µm	N	Y	~2.2 MB	NA	NA
C. A. arthropodicus	Rod 2–5 × 0.3 µm	Y	Y	~3.5 MB	Chromosome + small plasmid + 1 or more large plasmids	NA
C. Phlomobacter fragariae	Rod 2–2.5 × 0.3 µm	NA	NA	NA	NA	Phytopathogen
C. Riesia pediculicola	Rod Up to 25 µm long	N	N	0.57 MB	Linear chromosome + small plasmid	Obligate mutualist, B vitamin provision

Note: For references, see text. NA, not ascertained.

Around 5% of *Nasonia vitripennis* females are naturally infected with *A. nasoniae*, an infection otherwise known as son killer. The label *son killer* describes the phenotype of the infection. As discussed in the opening paragraph, *Arsenophonus* distorts the sex ratio of its wasp host *Nasonia*, increasing the proportion of daughters by killing up to 90% of sons (Skinner 1985; Werren et al. 1986).

The son-killer trait raises two questions: How does selective killing of sons occur, and why? The mechanistic question has been resolved in terms of the changes that occur to the host. As in other Hymenoptera, the wasp has haplodiploid sex determination: unfertilized (haploid) eggs develop into males, and fertilized (diploid) eggs develop into females. Indeed, *Arsenophonus nasoniae* appears to act by preventing the development of unfertilized eggs. Recently, *A. nasoniae* was shown to kill sons by preventing the formation of maternal centrosomes in unfertilized eggs, which resulted in early developmental arrest (Ferree et al. 2008). What is not clear mechanistically is the nature of *A. nasoniae* effectors that produce this change.

The advantage to the bacterium of son killing is less clear in this system. The general rationale for son killing is given in terms of benefit to a maternally inherited element that cannot pass through sons. Male killing may either directly release resources to female sibling hosts (e.g., cannibalism in ladybirds), release resources indirectly (through reduced resource competition following brood reduction), or reduce the probability of deleterious inbreeding (Hurst and Majerus 1993). *Nasonia* is a gregarious parasite that lays many eggs within a pupa. Death of males may therefore reduce resource competition suffered by female siblings within the pupa. *Nasonia* is also a species that routinely inbreeds, and this inbreeding avoidance has been suggested to provide a benefit to male killing in this system (Werren 1987). Experimental analysis of these factors has not been completed. With respect to the former hypothesis, field data indicated no difference in body size between infected and uninfected wasps (Balas et al. 1996). With respect to the latter, inbreeding occurs regularly in the species. However, inbreeding depression is modest (Luna and Hawkins 2004), and death of males may also result on occasion in females remaining virgin, which prevents the transmission of a maternally inherited element.

There are reasons to believe *A. nasoniae* will be biologically and microbiologically very different from other microbes that display reproductive parasitism. While *Arsenophonus* is a maternally inherited bacterium like *Wolbachia*, it is also horizontally transmitted at high frequency among *Nasonia* wasps developing within the same fly host. The initial observation of this predates the description of the bacterium, with Skinner (1985) noting transfer of the son-killer trait following superparasitism, and Huger et al. (1985) observing bacterial passage across gut epithelia. *Arsenophonus nasoniae* is thus inoculated into the fly host and then ingested by the developing wasp offspring. This peroral transmission of the bacterium to the next generation of wasps is unique, in contrast to the pure cytoplasmic mode of transmission typical of other sex ratio-distorting microorganisms. Unlike other reproductive parasites, *A. nasoniae* can be grown in cell-free media (Werren et al. 1986), allowing greater chance of survival outside the host that could enhance the likelihood of successful transfer between arthropods.

Horizontal transmission of heritable bacteria via ecological interactions has been proposed but has only rarely been demonstrated. Following its initial description in *N. vitripennis*, later studies also indicated *A. nasoniae* presence in *N. longicornis* (Balas et al. 1996). A wider survey of the filth fly community revealed *A. nasoniae* infection in two further parasite species, and in the bird nest fly *Protocalliphora azurea* (Diptera: Calliphoridae) itself (Duron et al. 2010). The infections were, in all cases, MLST identical to the type strain, suggesting the infection jumped recently between the different species in the filth fly community. Laboratory experiments were then used to test the hypothesis that *A. nasoniae* infection would transmit between species following multiparasitism (sharing of a pupal host), in the same way that intraspecific transmission of infection was observed. Host sharing was observed to lead to transfer of infection between species commonly, making *A. nasoniae* unusual among inherited symbionts in the degree to which it passes freely from species to species, certainly at higher rates than is regarded as normal for inherited bacteria (Duron et al. 2010).

Arsenophonus Arthropodicus/Hippoboscid Flies

Microbial Facts

Rod-shaped cells, measuring 2–5 μm in length and 0.3 μm in diameter. Possesses a genome of c. 3.5 MB with one small plasmid of 9.9 kb, and one or more larger plasmids. Culturable on cell-free liquid and solid MM media in a high CO_2/low O_2 atmosphere, where it grows slowly, producing visible colonies in 5 days, and colonies 2–3 mm diameter by 10 days. *In vitro* growth conducted at 25°C in the presence of polymyxin to prevent growth of other bacteria. Can be transformed with a broad host range plasmid (Dale et al. 2006).

Interaction with Host

Widely disseminated infection found throughout host tissues, and is found both intra- and intercellularly. Does not distort host sex ratio.

The blood-feeding flies of Hippoboscidae and Streblidae families harbor a high diversity of *Arsenophonus* strains (Dale et al. 2006; Duron et al. 2008; Novakova et al. 2009; Trowbridge et al. 2006). Two phylogenetically distinct strains are found in these families, but only one of these, Cand. *A. arthropodicus*, has been formally named. *Arsenophonus arthropodicus* is found in the pigeon louse fly, *Pseudolynchia canariensis* (Dale et al. 2006).

Very little is known about the evolutionary ecology of these *Arsenophonus* infections. Because *Arsenophonus* depend on maternal transmission for spread within arthropod populations, persistence of infection in a host species over long periods of time should result in diversification of symbiont alongside the host, the process of cocladogenesis. In accordance with this hypothesis, phylogenetic studies based on 16S rRNA gene sequences have suggested the codiversification of *Arsenophonus* with their hosts: monophyletic groups of *Arsenophonus* have been reported in flies of the families Hippoboscidae and Streblidae (Trowbridge et al. 2006). These results suggested that *Arsenophonus* acquisition could be ancient in these dipteran families,

and followed by vertical transmission that tracks host cladogenesis. However, this interpretation is based on 16S rRNA data and awaits confirmation from analysis of other genetic markers.

The phenotype of infections remains to be characterized. The *Arsenophonus* strains isolated from two different Hippoboscidae species show no evidence of sex ratio distortion activity (Dale et al. 2006; Duron et al. 2008): infection is equally present in males and female hosts, in contrast with the behavior of *A. nasoniae* in wasps. Obligate blood-sucking insects, such as the hippoboscids and streblids, live on a diet depauperate in vitamins, and other obligate blood feeders (e.g., lice, tsetse flies) obtain these vitamins from their symbionts. This represents a tempting area of study.

Arsenophonus Triatominarum/Triatomine Bugs

Microbial Facts

Highly filamentous rods, >15 μm in length, 1–1.5 μm in diameter. Cultivated in *Ae. albopictus* cell line C6/36 in MM medium supplemented with fetal calf serum at 25°C, where growth is visible in 72 h. Cryopreservation not yet successfully achieved, and culture on GC medium for *A. nasoniae* failed (Hypsa and Dale 1997). Preliminary work indicates a genome c. 2.2 MB in size. This modestly reduced genome indicates that cell-free culture will be challenging.

Interaction with Host

Infection widely disseminated in host. Bacteria reside intracellularly and extracellularly. Does not distort the sex ratio (Hypsa 1993).

As long ago as 1986, light and electron microscopy surveys of the salivary glands of two triatomine species, *Triatoma infestans* and *Panstrongylus megistus*, revealed the presence of intracellular infections (Louis et al. 1986). Later, in a more detailed study, the bacterium was found to possess a strict tissue tropism and to infect several tissues of the triatominae host. Most typically, and like *A. nasoniae*, infection is concentrated in neural ganglia and in large nests below the neurilemma. Visceral muscles, dorsal vessels, tracheal systems, gonadal sheaths, and hemocytes represent other tissues often invaded by the symbiont. Some tissues, such as somatic muscles or adipocytes in the fat body, were never found to contain the symbiont. Interestingly, developing ovarioles seem to be symbiont-free. However, the presence of symbionts in the embryonal gut indicates they enter the egg at some later stage of oogenesis/embryogenesis (Hypsa 1993).

Phylogenetic analysis of the 16S rRNA gene placed this bacterium as a sister group of the then phylogenetically isolated symbiotic bacterium, *Arsenophonus nasoniae*, which led to its description as a new *Arsenophonus* species, Cand. *A. triatominarum* (Hypsa and Dale 1997). Subsequently, *A. triatominarum* has been reported from other triatomine bugs and is currently known from 17 species (Sorfova et al. 2008). It thus represents the most numerous set of *Arsenophonus* lineages obtained from closely related hosts, making it an ideal system for the study of various aspects of the insect-*Arsenophonus* coevolutionary process, such as the age of

the association or the mode of host-symbiont cospeciation. In contrast to some other *Arsenophonus* lineages, *A. triatominarum* seems to have extremely high prevalence within the host population. To date, symbiont infection was detected in all individuals sampled over 17 investigated species of the tribe Triatomini. Its distribution covers the whole phylogenetic span of the tribe; it was detected in basal taxa, such as *Triatoma rubrofasciata*, in the most derived species, e.g., *T. infestans*, and also in several other genera (*Mepraia*, *Eratyrus*, *Panstrongylus*, *Meccus* (Sorfova et al. 2008)). On the other hand, it was not present in any of four tested species of the tribe Rhodniini (unpublished results). This pattern suggests a pronounced host specificity of this *Arsenophonus* lineage.

The coevolutionary history of host and symbiont is as yet unclear. Phylogenetic reconstruction established the strains of *A. triatominarum* as a monophyletic group. However, the phylogenetic arrangement of the *A. triatominarum* lineages does not correspond to the phylogeny of the host species (Sorfova et al. 2008). Since the reciprocal monophyly of host and symbiont clades is difficult to explain without invoking an ad hoc hypothesis (e.g., physiological constraints preventing infection of other host taxa), a coevolutionary history seems to be the most plausible explanation. Sorfova et al. (2008) postulated that the incongruence of host and symbiont phylogenies could be a phylogenetic artifact arising from intragenomic variability of the 16S rRNA gene sequence. Comparing two different copies of the 16S rRNA gene from *Arsenophonus* lineages isolated from four triatominae species, they showed that at this phylogenetic level, intragenome variability is capable of masking true phylogenetic relationships. The sequence of other genetic markers will be needed to resolve the issue.

There are no data available on the nature of the symbiosis between *A. triatominarum* and its host. Aposymbiotic bugs derived by antibiotic treatment remain viable and capable of reproduction (unpublished results). Patterns of molecular evolution and genome degeneration of *A. triatominarum* corresponds to those typical of S-symbionts rather than mutualistic P-symbionts. Compared to the presumably mutualistic long-branched lineages of *Arsenophonus* (i.e., *Riesia* and *Trichobius* symbiont), *A. triatominarum* displays only the standard baseline rate of 16S rRNA gene sequence evolution, and thus forms typical short-branched offshoots within the *Arsenophonus* phylogeny (Novakova et al. 2009). Other traits that make *A. triatominarum* distinct from mutualistic P-symbionts include only modest AT compositional shifts in housekeeping genes and the ability to maintain infection in cell culture.

The absence of any apparent effect of *A. triatominarum* on host fitness or reproduction raises the question of the factors maintaining this bacterium in the host population. Apart from beneficial effects typical for mutualistic symbionts, the most efficient way for a symbiont to spread through a host population is by manipulating host reproduction. However, no evidence of reproductive abnormalities has been observed in our laboratory colony of *T. infestans* (Hypša, personal observation). Furthermore, data obtained from the best studied model of reproductive manipulation, the genus *Wolbachia*, indicates that such manipulation only allows for a transient fixation of the symbiont in the host population, which is in contrast to the radiation of *A. triatominarum* in triatomine bugs.

Riesia Pediculicola/Lice

Microbial Facts

Filamentous rods, whose size varies through host development, with cells up to 25 μm in length (Perotti 2007). Possesses a genome of 0.574 MB with a single linear chromosome, and one small plasmid of 7.6 kb. Uncultured to date, a highly degenerate genome lacking an intact ATP synthesis pathway suggests microbe will be refractory to culture.

Interaction with Host

Found intracellularly within various mycetome structures, with location varying with host age. Extracellular migration between these structures, and to the ovariole (Perotti 2007). Effect on sex ratio unknown, although lice do produce female-biased broods (Perotti et al. 2004). Infection likely to provide host with B vitamins.

Human body and head lice (*Pediculus humanus*, *P. capitis*), human pubic lice (*Phthirus pubis*), and chimpanzee lice (*Pediculus schaeffi*) are all obligate blood-sucking parasitic insects that carry a required P-symbiont, located classically in stomach discs. Recently, the infection in each of these louse species was identified as belonging to a monophyletic bacterial group, called *Riesia*, which was allied to the genus *Arsenophonus*, with the infection of the three *Pediculus* species being monophyletic with respect to the infection found in pubic lice (Sasaki-Fukatsu et al. 2006; Allen et al. 2007). *Riesia* demonstrates cocladogenesis with its lice hosts, indicating that this is an obligate P-symbiont that has coevolved with these lice over their recent evolutionary past (Allen et al. 2007). All *Riesia* infections are required for host function, putatively providing B vitamins classically lacking in the diet of obligate blood feeders.

The clade is considered to have diverged in the last 13–25 million years, with the 16S rRNA gene in the clade *Riesia* showing the most rapid evolution of that found for this gene in any eubacteria, estimated at 19–34%/50 million years (Allen et al. 2009). The latter diversification contrasts with that of *Buchnera*, estimated at 1%/50 million years (Moran et al. 1995). The fast evolutionary rate is also reflected in the degeneration of the *Riesia* genome (see section below).

Despite the louse mycetome being the first ever discovered (observed by Hook in 1664 and Swammerdam in 1669), the biology of its mycetome symbiont *Riesia* has only recently been investigated fully (Perotti et al. 2007). Infection resides in two mycetomes that have prolonged existence across the louse life history, and move compartments in two transitory mycetomes, with two migration events where the bacteria exited the mycetome to move to either another mycetome or ovarian tissue. When within cells, the bacteria reside inside vacuoles. Outside cells, they are faced by a hostile cellular immune response, with hemocytes described as chasing the bacteria as they move from stomach disc to penetrate the oviduct, and phagocytosing those they encounter. The strength of the pursuit varies, occurring instantaneously on *Riesia* exiting cells in the case of head lice, but being delayed by 2 h in the case of body lice. It is conjectured that the structure of the tunica of the oviduct is an adaptation to allow ready penetration of the bacteria, and persistence of the symbiont despite an active cellular immune response.

Phlomobacter Fragariae/Planthoppers/Plants

Microbial Features

Rod-shaped microbe, 2–2.5 μm in length and 0.3 μm diameter. No culture or genomic information is currently available.

Interaction with Host

Microbe is largely associated with phloem of plants in which it causes disease, and is vectored by planthoppers. Infection in planthoppers is disseminated, including gut, salivary glands, and ovaries. Transmission electron microscopy (TEM) studies of infected insects indicate intracellular infection (Bressan et al. 2009). Effects on insect are unknown.

Arsenophonus strains (known as *Phlomobacter* spp.) have been found in planthoppers (Cixiidae) and also in the phloem of diverse plant species on which these insects feed, where they have been identified as the causative agents of phytopathologies: the syndrome "basses richesses" of sugar beet and the marginal chlorosis of strawberry (Bressan et al. 2008, 2009; Danet et al. 2003; Semetey et al. 2007; Zreik et al. 1998). Notably, *Arsenophonus* reduced the biomass and sugar content of sugar beet plants. In contrast, TEM study suggests infection of planthoppers with *Arsenophonus* is not associated with significant cytopathology.

The *Phlomobacter*-planthopper-plant interaction has obvious similarities to other phytopathogenic bacteria, where hemipteran hosts vector the bacteria between plants. This pattern is seen for *Spiroplasma* species (e.g., *S. kunkelli*, *S. citri*), as well as certain *Rickettsia* (e.g., the etelogic agent of papaya bunchy top disease (Davis et al. 1998)). For *Phlomobacter*, vertical transmission through the insect host occurs very inefficiently (30% of F1 progeny of infected mothers inherit infection (Bressan 2009)). Reflecting the two modes of transmission, qPCR indicates the two most heavily infected tissues of adult female planthoppers to be salivary glands and reproductive organs. Horizontal transmission through plants appears to be an epidemiologically dominant factor, with vertical transmission (and any potential benefit of infection) playing a secondary role in maintenance of the infection in the population (although it may be crucial in maintaining infection over winter in the absence of plants).

Aside from the cixiid bugs infected with *Arsenophonus* above, hemipterans in general appear to be a hotspot for infection. At least three other families of sap-feeding bugs host *Arsenophonus*: Aleyrodidae (whiteflies), Aphididae (aphids), and Psylloidea (psyllids) (Table 12.1). In the light of the *Phlomobacter* study, the role of plants in movement of these other *Arsenophonus* infections between their arthropod hosts is worthy of investigation. On the one hand, it is likely that some emergent phytopathologies will be linked to *Arsenophonus* infection, and a thorough examination of infection in hemipterans and their associated plants is now required. On the other hand, not all *Arsenophonus* infections in insects will have a biology like *Phlomobacter*. Indeed, recent investigations in whiteflies have found *Arsenophonus* as obligatory symbionts in bacteriocytes, i.e., inside host cells specifically modified to house bacteria (Gottlieb 2008). This location typifies infections that could pro-

vide an advantage to their hosts, although *Arsenophonus* could just benefit from the protection offered by the bacteriocyte.

ARSENOPHONUS GENOMES

Draft genome sequences have recently been completed for *A. nasoniae* (Darby et al. 2010) and *Riesia pediculicola* (Kirkness et al. 2010). Sequencing projects are under way for *A. triatominarum* and *A. arthropodicus.*

The Genome of *A. nasoniae*

As well as a reduced size relative to free-living relatives, the *A. nasoniae* genome also displays an AT bias. This conforms to the patterns seen in other vertically transmitted bacteria, despite *A. nasoniae*'s high levels of horizontal spread and its ability to infect across the gut epithelium. In terms of gene content, *A. nasoniae* shows no retention of metabolic pathways that obviously function in the provisioning of nutritional supplements to the host, as typify obligate symbiont genomes (Ruby 2008 from Darby et al. 2010). As might be expected of a bacterium that can be grown on supplemented cell-free media, *A. nasoniae* has abundant active metabolic pathways when compared to obligate symbionts such as *Buchnera* or *Wigglesworthia*, but shows a paucity of active biochemical pathways compared to free-living *E. coli* or *P. luminescens.* Analysis of the genome indicates that chitin may represent an important substrate for growth, which would accord with the saprophytic stage this bacterium has in the fly cadaver.

One feature of particular interest in the *A. nasoniae* genome is evidence of lateral transfer of genes from other symbionts. The similarity of phage elements in *Arsenophonus*, *Sodalis*, and *Hamiltonella* is one example of this. However, the presence of an open reading frame (ORF) encoding an outer membrane protein of *Wolbachia* in *A. nasoniae* indicates lateral transfer in the symbiome beyond the gamma-Proteobacteria (Darby et al. 2010). This suggests that mechanisms of symbiosis might readily move between quite unrelated symbionts.

A. nasoniae does, however, display an extensive and varied array of virulence and symbiosis factors (Wilkes et al. 2010). Notable among the virulence factors are four apoptosis-inducing protein (Aip)-like ORFs and several ORFs with an unusual chimeric structure, with an N terminus carrying multiple leucine-rich repeats (known to interact with ligands), and a C terminus carrying RTX and toxB domains. Rather than being used as host-lethal toxins, these ORFs are believed to play a role in host immune avoidance in *A. nasoniae*, probably secreted as a defense to phagocytosis. Other than this, the genome contains two type III secretion systems (TTSSs) and numerous ORFs with similarity to TTSS effectors of *Salmonella* sp. and *Yersinia* sp., as well as various other transport, adhesion, and flagellar apparatus. It is highly likely that these effectors are involved with disabling host innate immunity, and possibly in invasion through the gut. As a whole, it is clear that symbiosis and pathogenesis utilize similar mechanisms—aside from larger genome features, it is hard to discriminate between pathogen and symbiont in terms of the virulence genome.

The presence of regions with similarity to insecticidal genes of various gamma-Proteobacteria, but displaying significant pseudogenization, suggests *A. nasoniae* may have evolved from a generalist insect pathogen. The pseudogenized regions display similarity to the toxin complex (TC) genes of *Photorhabdus*, the repeats in toxin (RTX) genes of *Yersinia* and *E. coli*, and a pathogenicity island with similarity to virulence factors of *Xenorhabdus*, *Clostridium*, and *Ricketsiella* sp..

Despite a large toxin arsenal, the genetic basis of the male-killing phenotype of *A. nasoniae* remains elusive. Work by Ferree et al. (2008) identified male death to be the result of a lack of maternal centrosome formation in developing male (unfertilized) eggs. Candidate effectors are therefore either TTSS-injected or small, membrane-diffusible molecules. The presence of a polyketide synthase system (PKS) represents a possible solution, but further investigation is needed.

Genomic data place *A. nasoniae* at an intermediate stage in the evolution from free-living, insect-associated bacteria to obligate endosymbiont. Future work on the comparative genomics between *A. nasoniae* and other, non-male-killing *Arsenophonus* species will help elucidate the symbiosis set of genes used by all *Arsenophonus* species, and may further identify those factors involved in reproductive parasitism.

The Genome of *R. pediculicola*

The genome of *Riesia* contains just 557 ORFs, which reflects a typical highly reduced primary endosymbiont genome (Kirkness et al. 2010). Like other primary endosymbionts, it lacks mobile elements. This level of genome degradation is thought to have occurred very rapidly, the origin of *Riesia* in lice lying within the last 13–15 million years (Allen et al. 2009). The linear chromosome is notable, and is unique to date among endosymbionts. This chromosome possesses subtelomeric repeat elements that presumably provide stability against chromosomal erosion.

Comparative genomic analysis found *Riesia* to possess 24 ORFs not present in other obligate symbionts (but found in other bacteria). The majority of these genes are thought to be associated with transport and binding, and enzymes involved in lipopolysaccharide biosynthesis. This latter feature is conjectured as being important in the extracellular phase of *Riesia* life history, where it migrates from mycetome to ovariole. *Riesia* lacks exonuclease genes required for conjugation, and also many genes for ATP synthesis. The conjectured function of *Riesia*, provision of B vitamins to a host who lives on a depauperate blood diet, is reflected in an intact pathway for synthesis of vitamin B5 (pantothenic acid) split between chromosome and plasmid.

PROSPECTS

Members of the *Arsenophonus* genus have been observed to be common and likely to have a variety of interactions with their hosts. As noted above, while the type species *A. nasoniae* is a reproductive parasite, other members of the genus are likely secondary symbionts that could be conditionally beneficial, or primary symbionts, or insect-vectored phytopathogens. This mix of effects makes evolution in the clade *Arsenophonus* rather interesting, as there clearly are transitions between

reproductive parasitism and potentially beneficial symbiosis. The variety of phenotypes, once recapitulated to genome sequences, should provide an excellent place in which to generate hypotheses with respect to genes that function in different symbiotic capacities.

Aside from variation in the symbiotic interaction, *Arsenophonus* also provides a more tractable model for functional genetic research than most symbionts. It is culturable and clonable. It has been transformed with broad host range plasmids. It is likely that *Arsenophonus* will be a system in which the standard microbial toolkit can be adapted to produce GFP (green fluorescent protein)-marked strains, GFP-tagged proteins, as well as targeted knockouts. These can then be used for examining symbiont properties within hosts, for examining the targeting of proteins, and for direct testing of gene function.

We can identify four main future objectives for research on this symbiont:

1. Use of the recently developed multilocus typing system to understand how *Arsenophonus*-host interactions have evolved with respect to cospeciation and horizontal transfer, and any patterns in this.
2. Gaining a deeper understanding of the consequences of infection for the host. Currently, this is known only for *A. nasoniae* son killing and for *Riesia* anabolic roles. The effect of infection on other hosts is not known.
3. Further to this, the development of the existing systems above through genetic manipulation, in order to test the role of particular genes and pathways in establishing symbiosis and in symbiont phenotype.
4. The development of other *Arsenophonus*-host interactions for study. The five interactions detailed to date are likely a subset of the full range.

REFERENCES

Allen, J. M., Light, J. E., Perotti, M. A., Braig, H. R., and Reed, D. L. 2009. Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet. *PLoS ONE*, 4, e4969.

Allen, J. M., Reed, D. L., Perotti, M. A., and Braig, H. R. 2007. Evolutionary relationships of "*Candidatus* Riesia spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Appl. Environ. Microbiol.*, 73, 1659–1664.

Babendreier, D., Joller, D., Romeis, J., Bigler, F., and Widmer, F. 2007. Bacterial community structures in honeybee intestines and their response to two insecticidal proteins. *FEMS Microbiol. Ecol.*, 59, 600–610.

Balas, M. T., Lee, M. H., and Werren, J. H. 1996. Distribution and fitness effects of the son-killer bacterium in *Nasonia*. *Evol. Ecol.*, 10, 593–607.

Bressan, A., Semetey, O., Arneodo, J., Lherminier, J., and Boudon-Padieu, E. 2009. Vector transmission of a plant-pathogenic bacterium in the *Arsenophonus* clade sharing ecological traits with facultative insect endosymbionts. *Phytopathology*, 99, 1289–1296.

Bressan, A., Semetey, O., Nusillard, B., Clair, D., and Boudon-Padieu, E. 2008. Insect vectors (Hemiptera: Cixiidae) and pathogens associated with the disease syndrome "basses richesses" of sugar beet in France. *Plant Dis.*, 92, 113–119.

Clay, K., Klyachko, O., Grindle, N., Civitello, D., Oleske, D., and Fuqua, C. 2008. Microbial communities and interactions in the lone star tick, *Amblyomma americanum*. *Mol. Ecol.*, 17, 4371–4381.

Dale, C., Beeton, M., Harbison, C., Jones, T., and Pontes, M. 2006. Isolation, pure culture, and characterization of "*Candidatus* Arsenophonus arthropodicus," an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. *Appl. Environ. Microbiol.*, 72, 2997–3004.

Danet, J. L., Foissac, X., Zreik, L., Salar, P., Verdin, E., Nourrisseau, J. G., and Garnier, M. 2003. "*Candidatus* Phlomobacter fragariae" is the prevalent agent of marginal chlorosis of strawberry in French production fields and is transmitted by the planthopper *Cixius wagneri* (China). *Phytopathology*, 93, 644–649.

Darby, A. C., Choi, J. H., Wilkes, T., Hughes, M. A., Werren, J. H., Hurst, G. D. D., and Colbourne, J. K. 2010. Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*. *Insect Mol. Biol.*, 19, 75–89.

Davis, M. J., Ying, Z., Brunner, B. R., Pantoja, A., and Ferwerda, F. H. 1998. Rickettsial relative associated with papaya bunchy top disease. *Curr. Microbiol.*, 36, 80–84.

Dergousoff, S. J., and Chilton, N. B. 2010. Detection of a new *Arsenophonus*-type bacterium in Canadian populations of the Rocky Mountain wood tick, *Dermacentor andersoni*. *Exp. Appl. Acarol.*, 52, 85–91.

Dunn, A. K., and Stabb, E. V. 2005. Culture-independent characterization of the microbiota of the ant lion *Myrmeleon mobilis* (Neuroptera: Myrmeleontidae). *Appl. Environ. Microbiol.*, 71, 8784–8794.

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., and Hurst, G. D. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.*, 6, 27.

Duron, O., Wilkes, T. E., and Hurst, G. D. D. 2010. Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecol. Lett.*, 13, 1139–1148.

Ferree, P. M., Avery, A., Azpurua, J., Wilkes, T., and Werren, J. H. 2008. A bacterium targets maternally inherited centrosomes to kill males in *Nasonia*. *Curr. Biol.*, 18, 1409–1414.

Gherna, R. L., Werren, J. H., Weisburg, W., Cote, R., Woese, C. R., Mandelco, L., and Brenner, D. J. 1991. *Arsenophonus nasoniae* gen.-nov., sp.-nov., the causative agent of the son killer trait in the parasitic wasp *Nasonia vitripennis*. *Int. J. Syst. Bact.*, 41, 563–565.

Gottlieb, Y., Ghanim, M., Gueguen, G., Kontsedalov, S., Vavre, F., Fleury, F., and Zchori-Fein, E. 2008. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *FASEB J.*, 22, 2591–2599.

Grindle, N., Tyner, J. J., Clay, K., and Fuqua, C. 2003. Identification of *Arsenophonus*-type bacteria from the dog tick *Dermacentor variabilis*. *J. Invertebr. Pathol.*, 83, 264–266.

Hansen, A. K., Jeong, G., Paine, T. D., and Stouthamer, R. 2007. Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. *Appl. Environ. Microbiol.*, 73, 7531–7535.

Huger, A. M., Skinner, S. W., and Werren, J. H. 1985. Bacterial infections associated with the son-killer trait in the parasitoid wasp *Nasonia* (=Mormeniella) *vitripennis* (Hymenoptera Pteronalidae). *J. Invertebr. Pathol.*, 46, 272–280.

Hurst, G. D. D., and Majerus, M. E. N. 1993. Why do maternally inherited microorganisms kill males? *Heredity*, 71, 81–95.

Hypsa, V. 1993. Endoctytobionts of *Triatoma infestans*: distribution and transmission. *J. Invertebr. Pathol.*, 61, 32–38.

Hypsa, V., and Dale, C. 1997. *In vitro* culture and phylogenetic analysis of "Candidatus *Arsenophonus triatominarum*," an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int. J. Syst. Bact.*, 47, 1140–1144.

Kirkness, E. F., Haas, B. J., Sun, W. L., Braig, H. R., Perotti, M. A., Clark, J. M., Lee, S. H., Robertson, H. M., Kennedy, R. C., Elhaik, E., Gerlach, D., Kriventseva, E. V., Elsik, C. G., Graur, D., Hill, C. A., Veenstra, J. A., Walenz, B., Tubio, J. M. C., Ribeiro, J. M. C., Rozas, J., Johnston, J. S., Reese, J. T., Popadic, A., Tojo, M., Raoult, D., Reed, D. L., Tomoyasu, Y., Krause, E., Mittapalli, O., Margam, V. M., Li, H. M., Meyer, J. M.,

Johnson, R. M., Romero-Severson, J., Vanzee, J. P., Alvarez-Ponce, D., Vieira, F. G., Aguade, M., Guirao-Rico, S., Anzola, J. M., Yoon, K. S., Strycharz, J. P., Unger, M. F., Christley, S., Lobo, N. F., Seufferheld, M. J., Wang, N. K., Dasch, G. A., Struchiner, C. J., Madey, G., Hannick, L. I., Bidwell, S., Joardar, V., Caler, E., Shao, R. F., Barker, S. C., Cameron, S., Bruggner, R. V., Regier, A., Johnson, J., Viswanathan, L., Utterback, T. R., Sutton, G. G., Lawson, D., Waterhouse, R. M., Venter, J. C., Strausberg, R. L., Berenbaum, M. R., Collins, F. H., Zdobnov, E. M., and Pittendrigh, B. R. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. USA*, 107, 12168–12173.

Louis, C., Drif, L., and Vago, C. 1986. Evidence and ultrastructural-study of *Rickettsia*-like prokaryotes in salivary-glands of Heteroptera-Triatomidae. *Ann. Soc. Entomol. France*, 22, 153–162.

Luna, M. G., and Hawkins, B. A. 2004. Effects of inbreeding versus outbreeding in *Nasonia vitripennis* (Hymenoptera: Pteromalidae). *Environ. Entomol.*, 33, 765–775.

Moran, N. A., Mccutcheon, J. P., and Nakabachi, A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.*, 42, 165–190.

Moran, N. A., Von Dohlen, C. D., and Baumann, P. 1995. Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *J. Mol. Evol.*, 41, 727–731.

Najar-Rodriguez, A. L., Mcgraw, E. A., Mensah, R. K., Pittman, G. W., and Walter, G. H. 2009. The microbial flora of *Aphis gossypii*: patterns across host plants and geographical space. *J. Invertebr. Pathol.*, 100, 123–126.

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

Perotti, M. A., Allen, J. M., Reed, D. L., and Braig, H. R. 2007. Host-symbiont interactions of the primary endosymbiont of human head and body lice. *FASEB J.*, 21, 1058–1066.

Perotti, M. A., Catala, S. S., Ormeno, A. D., Zelazowska, M., Bilinski, S. M., and Braig, H. R. 2004. The sex ratio distortion in the human head louse is conserved over time. *BMC Genet.*, 5.

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

Sasaki-Fukatsu, K., Koga, R., Nikoh, N., Yoshizawa, K., Kasai, S., Mihara, M., Kobayashi, M., Tomita, T., and Fukatsu, T. 2006. Symbiotic bacteria associated with stomach discs of human lice. *Appl. Environ. Microbiol.*, 72, 7349–7352.

Semetey, O., Gatineau, F., Bressan, A., and Boudon-Padieu, E. 2007. Characterization of a gamma-3 Proteobacteria responsible for the syndrome "basses richesses" of sugar beet transmitted by *Pentastiridius* sp. (Hemiptera, Cixiidae). *Phytopathology*, 97, 72–78.

Skinner, S. W. 1985. Son-killer: a third extrachromosomal factor affecting sex ratios in the parasitoid wasp *Nasonia vitripennis*. *Genetics*, 109, 745–754.

Sorfova, P., Skerikova, A., and Hypsa, V. 2008. An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of *Arsenophonus triatominarum*. *Syst. Appl. Microbiol.*, 31, 88–100.

Spaulding, A. W., and Von Dohlen, C. D. 2001. Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing mutation in ribosomal RNA. *Ins. Mol. Biol.*, 10, 57–67.

Subandiyah, S., Nikoh, N., Tsuyumu, S., Somowiyarjo, S., and Fukatsu, T. 2000. Complex endosymbiotic microbiota of the citrus psyllid *Diaphorina citri* (Homoptera: Psylloidea). *Zool. Sci.*, 17, 983–989.

Thao, M. L., Clark, M. A., Baumann, L., Brennan, E. B., Moran, N. A., and Baumann, P. 2000a. Secondary endosymbionts of psyllids have been acquired multiple times. *Curr. Microbiol.*, 41, 300–304.

Thao, M. L., Moran, N. A., Abbot, P., Brennan, E. B., Burckhardt, D. H., and Baumann, P. 2000b. Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Appl. Environ. Microbiol.*, 66, 2898–2905.

Thao, M. L. L., and Baumann, P. 2004. Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae). *Curr. Microbiol.*, 48, 140–144.

Trowbridge, R. E., Dittmar, K., and Whiting, M. F. 2006. Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *J. Invertebr. Pathol.*, 91, 64–68.

Wang, W.-X., Luo, J., Lai, F.-X., and Fu, Q. 2010. Identification and phylogenetic analysis of symbiotic bacteria *Arsenophonus* from the rice brown planthopper, *Nilaparvata lugens* (Stal) (Homoptera: Delphacidae). *Acta Entomol. Sin.*, 53, 647–654.

Werren, J. H. 1987. The coevolution of autosomal and cytoplasmic sex ratio factors. *J. Theor. Biol.*, 124, 317–334.

Werren, J. H. 2005. Arsenophonus. In Garrity, T. M. (ed.), *Bergey's manual of systematic bacteriology.* New York: Springer Verlag. p. 176.

Werren, J. H., Skinner, S. W., and Huger, A. M. 1986. Male-killing bacteria in a parasitic wasp. *Science*, 231, 990–992.

Wilkes, T. E., Darby, A. C., Choi, J. H., Colbourne, J. K., Werren, J. H., and Hurst, G. D. D. 2010. The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. *Insect Mol. Biol.*, 19, 59–73.

Zchori-Fein, E., and Brown, J. K. 2002. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Ann. Entomol. Soc. Am.*, 95, 711–718.

Zreik, L., Bove, J. M., and Garnier, M. 1998. Phylogenetic characterization of the bacterium-like organism associated with marginal chlorosis of strawberry and proposition of a *Candidatus* taxon for the organism, '*Candidatus* Phlomobacter fragariae.' *Int. J. Syst. Bacteriol.*, 48, 257–261.